DOCUMENT RESUME

ED 235 629	EC 160 486

AUTHOR	Holovet, Jennifer, Ed.; Guess, Doug, Ed.
TITLE	The KICS Model: Evaluation Studies.
INSTITUTION	Kansas Univ., Lawrence. Dept. of Special Education.; Topeka Unified School District 501, Kans.
SPONS AGENCY	Office of Special Education and Rehabilitative Services (ED), Washington, DC. Div. of Innovation and Development.
PUB DATE	83
NOTE	390p.; Pages 2-36 may not reproduce well.
PUB TYPE	Reports - Research/Technical (143) -- Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC16 Plus Postage.
DESCRIPTORS	Data Collection; *Drills (Practice); Elementary Secondary Education; *Generalization; Grouping (Instructional Purposes); Individualized Instruction; Models; Program Effectiveness; Program Evaluation; Recordkeeping; *Severe Disabilities
IDENTIFIERS	*Kansas Individualized Curriculum Sequencing

ABSTRACT

The manual presents 12 papers summarizing research on the effectiveness of the Kansas Individualized Curriculum Sequencing (KICS) model for severely handicapped students. The first three papers examine the effects of distributed practice schedules on the learning, generalization and initiation of students. The use of distributed practice, an important element in the KICS model, was supported by the three studies. Two succeeding papers explore the effects of the model on initiation and generalization. Little difference was found in the response initiation variable between KICS and massed trial training, while multiple exemplar training (as evidenced in KICS) was found to produce generalization across items requiring similar response topographics. The third set of studies supports the effectiveness of group instruction for students with severe and multiple needs. Studies featured structured interaction between students, examination of incidental learning, and effects of group arrangements on vocational workshop tasks. Data recording systems are addressed by three studies which emphasized the need for frequent, although not necessarily daily, recording. A final paper investigates the effects of allowing severely handicapped adolescents to make limited choices of educational activities. (CL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED235629

The KICS Model: Evaluation Studies

Edited by

Jennifer Holvoet

Doug Guess

With significant contributions from
Marilyn Mulligan
Ed Helmstetter
Nancy Schussler
Fredda Brown

1983

The Kansas Individualized Curriculum Sequencing Project,
University of Kansas
Department of Special Education
Lawrence, Kansas

EC 160 486

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Jennifer Holvoet
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Preparation of this manual was supported in part by the U.S. Department of Education, Office of Special Education, Division for Innovation and Development, awarded to the University of Kansas, Department of Special Education (Doug Guess, Principal Investigator). The opinions expressed herein, however, do not necessarily reflect the position or policy of the U.S. Department of Education, and no official endorsement by the Department should be inferred.

DISCRIMINATION PROHIBITED--No person in the United States shall, on the basis of race, color, or national origina, be excluded from participation in, be denied the benefits of, or be sujected to discrimination under any program or activity receiving Federal financial assistance, or be so treated on the basis of sex under most education programs or activities receiving Federal assistance.

TABLE OF CONTENTS

A. Evaluation of the Effects of Distributed Practice Trials 1.

1. Effects of massed, distributed and spaced trial sequencing on severely handicapped students' performance. 2.
   by M. Mulligan, L. Lacy, and D. Guess

2. The effects of number and composition of interspersed trials on the learning of severely handicapped students. 37.
   by J. Holvoet and L. Fernandez

3. Effects of distributed practice instruction on the cross-setting generalization of severely handicapped adolescents. 65.
   by D. Brewer, J. Holvoet, M. Mulligan & D. Carr

B. Evaluation of the Effects of the Individualized Curriculum Sequencing Model on Initiation and Generalization 91.

1. Initiation responses of persons with severe handicaps during Individualized Curriculum Sequence (ICS) training. 92.
   by L. Lacy, M. Mulligan, J. Holvoet, & D. Guess

2. Effects of multiple exemplar training on generalization of prevocational skills. 122.
   by B. Brown, J. Holvoet, & D. Guess

C. Evaluation of the Effects of Group Training 143.

1. The effect of systematic peer interaction on the incidental learning of two severely handicapped students. 144.
   by F. Brown and J. Holvoet

2. The impact of group instruction on incidental learning in learners with severe mental retardation. 170.
   by J. Rea, J. Holvoet, and N. Schussler

3. Effects of an assembly-line format on the work productivity of retarded students. 191.
   by J. Holvoet, B. Brown, N. Schussler, & D. Guess

D. Evaluation of Data Recording Systems 221.

1. How many data are enough? 222.
   by J. Holvoet, M. Mulligan, D. Guess & N. Schussler

2. Do we really have to take data? 241.
   by J. Holvoet, C. O'Neil, L. Chazdin, D. Carr, & J. Warner

3. The effects of two different data sheet formats on prompting accuracy. 275.
   by J. Holvoet, N. Schussler, L. Vogt, J. Rea, & L. O'Shea

E. Evaluation of Choice and Ready Responses on the Learning of Other skills 306.

1. Influence of activity choice on learning among adolescent students with severe handicaps. 307.
   by J. Holvoet, M. Brewer, M. Mulligan, D. Guess, E. Helmstetter, & P. Riggs

Preface

This manual summarizes research undertaken to evaluate the effectiveness of several components of the Kansas Individualized Curriculum Sequencing (KICS) model. A description of the theoretical bases of the model can be found in:

1. Guess, D., Horner, R. D., Utley, B., Holvoet, J., Maxon, D., Tucker, D., & Warren, S. The functional curriculum sequencing model for teaching the severely handicapped, AAESPH Review, 1978, 4(3), 202-215.
2. Mulligan, M., Guess, D., Holvoet, J., & Brown, F. The Individualized Curriculum Sequencing Model (I): Implications from research on massed, distributed or spaced trial training. Journal of the Association for the Severely Handicapped, 1980, 5(4), 325-336.
3. Holvoet, J., Guess, D., Mulligan, M., & Brown, F. The Individualized Curriculum Sequencing Model (II): A teaching strategy for severely handicapped students. Journal of the Association for the Severely Handicapped, 1980, 5(4), 337-351.
4. Brown, F., Holvoet, J., Guess, D., & Mulligan, M. The Individualized Curriculum Sequencing Model (III): Small group instruction. Journal of the Association for the Severely Handicapped, 1980, 5(4), 32-367.

A detailed explanation of how to implement the KICS model is contained in:

Holvoet, J., Mulligan, M., Schussler, N., Lacy, L., & Guess, D. The KICS Model: Sequencing learning experiences for severely handicapped children and youth. Lawrence, KS: University of Kansas, Department of Special Education, 1982.

This volume serves to evaluate the effectiveness of several of the components of the model that were described in the previous references. These evaluation efforts were developed, implemented, and refined through the efforts of several agencies and individuals. These agencies and individuals have contributed their ideas, direct services, students, encouragement and support throughout the processes of development, field-testing and evaluation of the model. Without them, no progress could have been made.

These evaluation efforts were a joint effort between the Department of Special Education at the University of Kansas and the Special Services Division of the Topeka Public School System (U.S.D. 501). The research and field-testing were, for the most part, conducted in self-contained U.S.D. 501 classrooms for severely handicapped adolescents. These could never have been successful without the help and cooperation of public school personnel. In particular, the efforts of Winn Green, Norma Blankinship, and the teachers and paraprofessionals who participated are appreciated. The input of teachers Perrin Riggs, Eileen Luddy, Sue Storms, and Liz Vogt was invaluable in making this evaluation effort focus on issues of interest to direct-service personnel.

Numerous persons from the Department of Special Education at the University of Kansas and from several departments at Washburn University in Topeka have given unstintingly of their time and expertise. These persons include: Nancy Schussler, Cheryl Watkins, Leslie Blankinship, Mike Brewer, George Warrick, Laurie O'Shea, Lynette Lacy, Mechthilde Heron, Sultana Aziz, Jerry Rea, Blane Brown, Robyn Potashnik, Ellen Mellard, and Cathy Neal. These were the people who did the day-to-day work during the field testing and evaluation.

Thanks also goes to several other people who were involved in the conceptualization of the research studies and in the review process. In particular, the contributions of Wayne Sailor, Bonnie Utley, Donna Lehr, Fredda Brown, Karen Barnes, Lesley Fernandez, Nancy Schussler, Marilyn Mulligan, and Ed Helmstetter were most helpful.

Lastly, we would like to extend our sincere appreciation to the students in the experimental classrooms and to their parents for all their cooperation in helping us to make this dream a reality.

Evaluation of the Effects of Distributed Practice

This section contains several studies which were undertaken to ascertain the effects of massed, spaced and distributed practice schedules on the learning, generalization and initiation of students with severely handicapped conditions. Distributed practice is the foundation of the Individualized Curriculum Sequencing Model.

The first study was a group study and looked at learning and generalization under three practice schedules. The second study used a single-subject design and investigated whether the way trials were distributed, or the number of distributed trials, had an effect on student learning.

The third study compared generalization under massed and distributed (ICS) conditions using a single-subject design.

All the studies indicated the usefulness of the distributed practice schedule particularly in reducing "refusals to respond."

Effects of Massed, Distributed and Spaced Trial Sequencing on Severely Handicapped Student' Performance[1]

Marilyn Mulligan, Ph.D.
University of Kansas
Lawrence, Kansas

Lynette Lacy, M.S. Ed.
U.M.K.C. UAF Preschool
Kansas City, Missouri

Doug Guess, Ph.D.
University of Kansas
Lawrence, Kansas

---

[1] This paper is reprinted with the permission from The Journal of the Association for the Severely Handicapped, 1982, 7(2), 48-61.

The authors wish to thank Lake Mary School, Paola, Kansas; USD 501, Topeka, Kansas; Brown Co. Special Education Cooperative, Hiawatha, Kansas for their cooperation in the implementation of this study. Funds for this study were provided through the U.S. Department of Education Student Research Fund, Grant No. 443 AH 0087.

ABSTRACT

The effects of massed, distributed, and spaced trial sequencing on performance during the training of cognitive and motor skills were evaluated. Eleven severely handicapped students in four isolated special education classrooms were involved. Three types of responses were monitored. These included correct, error, and refusal responses across the three conditions of training: massed, distributed, and spaced trial sequencing. An overall comparison of the three conditions of trial scheduling showed no difference in the mean level or the slope of the three types of responses. A comparison of distributed and massed trial sequencing only showed the superior effect of distributed trial sequencing for the mean level of responding. These findings were significant 1) for both correct and refusal responses for an overall comparison and 2) for students living in the home rather than in an institution. The results of this study support the addition of distributed trial sequencing to the technology of instruction for severely handicapped learners.

INTRODUCTION

The topic of this study is whether the technique of separating and distributing training trials for a particular skill across the entire school day should be incorporated into our teaching technology. The question grows out of a concern that the procedures for teaching severely handicapped students are not always maximizing instructional time, nor are these procedures attending to the particular learning characteristics that separate severely handicapped students from other, less handicapped learners. It is recognized that a slow acquisition rate and a reduced ability to generalize are both learning characteristics of the severely handicapped individual (Brown, Nietupski, & Hamre-Nietupski, 1976; Guess, Horner, Utley, Holvoet, Maxon, Tucker, & Warren, 1978). Some persons (Mulligan, Guess, Holvoet, and Brown, 1980; Holvoet, Guess, Mulligan, and Brown, 1980) have proposed the use of an Individualized Curriculum Sequence (ICS) as a technology to address these concerns. The most controversial element of this sequencing model is the procedure of placing training trials from one skill program between separate training trials from other skill development programs. This represents a significant departure from current methods of trial sequencing typically utilized in classrooms for severely handicapped learners.

The ICS is derived from the premise that the severely handicapped student needs to learn the interrelationships between skills being taught in conjunction with the acquisition of the skill itself (Holvoet & Utley, Note 1; Holvoet, Guess, Mulligan, & Brown, 1980). This technology involves sequencing trials for separate skills in a manner that emphasizes the logical or useful relation of one skill to another. For example, a student might orient towards one object (visual fixation), produce a sound in imitation that approximates the label of the object (expressive communication), reach for and grasp the object (fine motor), and then produce another sound in imitation that approximates the common action of the object (expressive communication).

The basis for this curriculum structure is the research evidence in the experimental psychology literature indicating that a procedure of distributing training trials may, in some cases and conditions, be superior to a procedure of massed trials training. This literature has been presented in a previous publication (Mulligan, Guess, Holvoet, and Brown, 1980) and will be briefly summarized here.

Operational Definition of Massed, Distributed, and Spaced Trials

The difference between the procedure of massing, distributing, or spacing trials essentially centers on what occupies the time between two trials from the same program. Given a period of instruction, trials from a program are massed if two repeated trials occur so closely together that no other behavior can be expected to be emitted between these two trials. Both spaced and distributed trial sequencing have separated trials, or periods of time occuring between two trials from the same program. Trials from a program are distributed if trial(s) from another program or programs occur between the two repeated trials from the same program. Trials are spaced if a rest period or pause occurs between two repeated trials from the same program (Underwood, Kapelak, & Malmi, 1976).

The Effects of Distributed, Spaced, and Massed Trials on Learning

Attention to the scheduling of training trials is based on the recognition that the learning of any particular skill is a function of the condition under which it is learned, regardless of the characteristics of the learner. According to Deese (1958), the effect of the separation of trials throughout time on learning has been explored more than any other condition of practice. Following the definitions presented earlier, results indicated that skills with a spaced or distributed trial sequence are learned better than skills learned using massed trials. This has been demonstrated primarily for nonfunctional cognitive and motor tasks both with handicapped and nonhandicapped individuals. The cognitive tasks used to demonstrate this effect are reading (Gargagliano, Note 2), learning nonsense words or syllables (Hovland, 1940a, 1940b), and paired associate learning of words (Dent & Johnson, 1964; Madsen, 1963; Underwood & Goad, 1951; Hovland, 1939). The motor tasks used are a selfhelp "shirt on" task and scanning task (Helmstetter, Note 3), gross motor stability (Chasey, 1975; Stelmach, 1969), fine motor rotary pursuit (Reynolds & Adams, 1953; Duncan, 1951; Adams, 1952; Kimble, 1949a), and fine motor manipulation of pegs (Kimble & Bilodeau, 1949; Carron, 1969). Some studies demonstrating the superiority of spaced versus massed trials have combined both motor and cognitive tasks by requesting subjects to draw lines between consecutive numbers randomly placed on a page (Baumeister & Berry, 1976; Underwood, 1961) or perform a printing task (Kimble, 1949a, 1949b; Kientzle, 1946).

An overall analysis of the learning curves comparing distributed, massed, and spaced trials schedules reveals two significant characteristics: 1) the spacing effect is not a linear relation; and 2) performance under these conditions of practice can be shifted from low to high, and high to low levels following a shift in the schedule of the trials. The spacing of trials primarily impacts on the middle portion of the learning curve. The first several trials under a spaced schedule are indistinguishable from those sequenced on a massed schedule (Kimble, 1949a). After initial learning, performance then becomes a negatively accelerating function (the curve climbs less dramatically over time) of both the length of the rest period (Kimble, 1949a; Kientzle, 1946) and the number of trials (Duncan, 1951; Adams, 1952). The general conclusion is that the overall effect of massing trials is a depression in performance during the asymptote or high point of the learning curve.

Additional factors seeming to affect performance under different conditions of practice are the similarity of the items being learned, criterion levels of responding, and the conspicuousness of the discriminative stimuli. Underwood and Goad (1951) taught lists of similar and dissimilar adjectives under conditions of spaced and massed practice. They found no difference in learning for dissimilar lists while there was a differential effect in favor of spaced trials for similar lists. Underwood, Kapelak, & Malmi (1976) found that the spacing effect occured when the subject had high or difficult criteria levels for responding, but this effect was not present for low criteria levels. Elmes, Sanders and Dovel (1973) demonstrated that when the cues for trials under the condition of mass practice were differentiated from ongoing activities (i.e. the use of different colors), performance was similar to the effect of spacing when the cues were not distinguishable from ongoing activities. However, when the cues for trials under both massed and spaced conditions were differentiated, the superiority of the spaced condition was maintained. These factors all relate to the difficulty of the skill being learned. The evidence seems to indicate the more unfamiliar or demanding the task, the more a distributed trials strategy assists performance.

The brief review presented here suggests that distributed trial training provides an additional approach to trials scheduling that is consistent with the ICS Model being proposed. Support is offered with the caution that most of the studies reviewed were conducted with non-handicapped subjects, with nonfunctional tasks, and often with learning content that was not appropriate to the skill level of severely handicapped students.

The present study applies the literature pertinent to this issue and investigates the overall efficacy question of including the use of distributed trial training in the technology of instruction for severely handicapped individuals.

METHOD

Subjects

Eleven subjects were selected from six participating classrooms for severely multiply handicapped students in four different locations according to the following criteria: severely and multiply handicapped (i. e., identified as having mental retardation with at least one additional severely handicapping condition); under instructional control for at least one response; and having programs targeted by the Individualized Educational Plan (IEP) in both motor and cognitive areas. The classroom teachers made the final selection of the students to participate in the study with respect to the availability of staff and student time.

Table 1

Secondary Handicapping Characteristics of Students in Addition to the Primary Handicapping Condition of Severe Mental Retardation

Student Number	Sex	Birth Date	Secondary Characteristics	Test/Approximate Functioning Level
1	F	6-61	Deaf/Blind	TARC/52%
2	M	11-66	Deaf/Blind	TARC/43%
3	F	7-66	Seizure disordered	Callier-Azusa/30 mos.
4	F	7-72	Seizure disordered	Callier-Azusa/16 mos.
5	M	12-62	Severe Cerebral Palsy Non-ambulatory	TARC/85% AAMD/45%
6	M	12-63	Severe Cerebral Palsy Non-ambulatory	TARC/49%
7	M	9-76	Severe Cerebral Palsy	Callier-Azusa/12 mos.
8	F	1-65	Autistic-like	TARC/67%
9	M	11-69	Developmentally Delayed Autistic-like	TARC/61%
10	F	9-73	Developmentally Delayed	Callier-Azusa/10 mos.
11	F	4-74	Developmentally Delayed	Callier-Azusa/16 mos.

The five boys and six girls participating ranged in age from five to nineteen years. Their combined characteristics covered most possible handicapping conditions from vision impairment to cerebral palsy, mental retardation, and seizure disorder. Table 1 represents the primary and secondary handicapping characteristics of the students and their approximate functioning level.

Insert Table I about here

Settings

All training occurred in the students' classroom or in an adjacent individual session room. The classrooms were typical of those serving severely handicapped students. Tables, chairs, adapted equipment, and age appropriate materials were used depending on the requirements of the programs for each student. Training occurred in the same location and at approximately the same time of day, three to five days per week. The same teacher conducted the sessions for each student throughout the study and utilized teaching techniques typical of those alread implemented in the classroom.

Both reliability observations and videotaping occurred during the the study. Attempts were made to make these extra observations as inconspicuous as possible, however no dividers or oneway mirrors were used.

Response Definition

Response Observation. Each teacher used one data sheet per session. The data sheet indicated the trials and the order in which the trials were to occur. The teacher recorded the performance for each student for each trial on the data sheet while conducting the session. Figure 1 presents a sample data sheet for one session.

Insert Figure 1 about here

Response Measure. Three major levels of responding were recorded: correct, incorrect, and refusal to respond. The actual coding of the response varied across teachers, students, and programs. A response was considered correct if it was independent, unprompted, or completed with one verbal instruction. A response was incorrect if the student received either a verbal or physical prompt or made an error. A response was considered a refusal if the student turned away, resisted, or did not participate in an approximation of the target response.

Procedures

The following procedures describe how training on both cognitive and motor programs was conducted across massed, spaced, or distributed training trial schedules.

Training programs. Each student was involved in four training programs throughout the study. Out of the four, at least one program was based on the development of a cognitive skill and one was based on the development of a motor skill.

Programs for each student were developed independently of the training programs for the other students. The programs were developed from the student's IEP with the cooperation of the teacher and the experimenter. The program plan followed the teacher's own curriculum development style and instructional techniques as closely as possible.

Data Sheet

Student No. 13
Name Jim
Teacher Tom
Condition Distributed

Date 10/17/80
Day 1
Time start 10:00 stop 10:36

Trial No.	Task	Score	Trial No.	Task	Score
1	Comm. Thirsty		21	Comm. Thirsty	
2	Folding		22	Folding	
3	Pict/Obj. Match Car		23	Pict/Obj. Brush	
4	Table Set		24	Table Set	
5	Comm. Thirsty		25	Comm. Thirsty	
6	Folding		26	Folding	
7	Pict/Obj. Brush		27	Pict/Obj. Car	
8	Table Set		28	Table Set	
9	Comm. Thirsty		29	Comm. Thirsty	
10	Folding		30	Folding	
11	Pict/Obj. Car		31	Pict/Obj. Brush	
12	Table Set		32	Table Set	
13	Comm. Thirsty		33	Comm. Thirsty	
14	Folding		34	Folding	
15	Pict/Obj. Car		35	Pict/Obj. Car	
16	Table Set		36	Table Set	
17	Comm. Thirsty		37	Comm. Thirsty	
18	Folding		38	Folding	
19	Pict/Obj. Brush		39	Pict/Obj. Brush	
20	Table Set		40	Table Set	

Figure 1. Sample data sheet indicating the order of training trials across four programs for training in the Distributed Condition.

Elements common across all programming styles were training procedure, correction procedure, task analysis (if applicable), and the data code used.

Modifications, or phase changes, in the training programs, were made when the student was performing at the specified criterion level or if performance was declining. Criterion levels were independently determined for each training program for each student by the classroom teacher. Typically the levels were specified between eighty and ninety percent of responses correct for three consecutive training days. Declining performance was also a judgment made by the classroom teacher. Usually this was reflected in a negative slope for four to five consecutive days. Phase changes for criterion responding required at least two days in each condition before the next phase of the program could be implemented. This required at least six sessions in a phase if the student was responding at maximum levels. Phase changes as a result of declining or flat trends during the acquisition period of training required at least three days in each condition before the phase change could be implemented. This was a minimum of nine sessions across the three experimental conditons. Phase changes included anything from modifications in the training techniques, the criterion response, behavior management techniques, materials used, or adaptive equipment for the student.

Training programs utilized during the study could also be used throughout the remainder of the school day. If a program was implemented during the remainder of the school day, it was consistently implemented throughout all conditions of training.

Training sessions. Each training session consisted of forty training trials, ten trials per training program, four training programs per session. The four training programs included at least one cognitive and one motor program, with the other two either in cognitive or motor domains. The data sheet for each day of training ordered the forty trials in either a massed, spaced, or distributed trials schedule. Sessions in the massed condition ordered the training trials so that all ten trials of one program were completed before the ten trials for the following programs were begun. Sessions in the spaced condition followed the same type of sequencing except there was a space of fifteen to twenty seconds between any two trials. Sessions with distributed trials ordered the training trials so that a cluster of four trials, one from each training program, was repeated ten times. This resulted in two trials from the same program separated by one trial from each of the other three programs.

The order of programs within each condition was initially randomly determined and then remained constant across the remainder of the study. Figure 2 is a Sample Schedule Sheet showing one student's randomizations schedule. "Daily Randomization" shows the condition occurring on each day of training and the ordering of the training trials within conditions.

Data sheets were developed separately for each session and followed the ordering of trials from the Schedule Sheet. Figures 1 and 3 are the data sheets used for the massed and distributed conditions indicated on the Sample Schedule Sheet (Figure 2). The numbers 2, 1, 3, 4 in the "Massed Condition" (Figure 3) indicate that ten trials of Program 2, Table Setting, occurred before ten trials of Program 1, Picture/Object

Match. These were followed by ten trials of Program 3, Folding Washclothes, then ten trials of Program 4, Communication. A different series of random numbers 4, 1, 2, 3, indicate the order of the forty training trials during the "Spaced Condition". The numbers 4, 3, 1, 2 for the "Distributed Condition" indicate that a set of four trials, one from each numbered program, occurred ten times in this order. For each training session, the teacher followed the order as indicated on the appropriate data sheet and scored the student's performance as it occurred.

Insert Figures 2 & 3 about here

Experimental Design

The experimental design consisted of three conditions; massed, spaced, and distributed trials scheduling across four training programs for each student. The conditions were randomly sequenced across training sessions following an Alternating Treatments Design (Barlow & Hayes, 1979). The previously presented, Figure 2 shows one subject's randomization of conditions. Each condition lasted a maximum of ten sessions. Due to the randomization procedure the number of sessions per condition did vary across students. The study was terminated for each student at the end of thirty training sessions or when the student reached criterion performance on three of the four training programs. Reliability

Reliability observations occurred between one and four times for each student, condition, and phase of training (if applicable). The reliability observer was present during the session and scored the

Schedule Sheet

Student No. 13

Student Jim

Location Sherridan

Teacher Tom

Program 1. Picture/Object Match
2. Table Setting
3. Folding washcloths
4. Communication - Thirsty

Daily Randomization; Distributed = 1, Massed = 2, Spaced = 3

Day	Schedule	Day	Schedule	Day	Schedule	Day	Schedule
1	Spaced	11	Massed	21	Massed	31	Spaced
2	Dist.	12	Massed	22	Spaced	32	Dist
*3	Spaced	*13	Dist.	*23	Massed	33	Spaced
4	Spaced	14	Dist.	24	Massed	34	Spaced
5	Dist.	15	Spaced	25	Dist.	35	Massed
6	Massed	16	Dist.	26	Massed	*36	Massed
7	Spaced	17	Massed	27	Massed	37	Spaced
*8	Spaced	18	Dist.	*28	Dist.	38	Massed
9	Dist.	19	Dist.	29	Massed	39	Spaced
10	Spaced	20	Massed	30	Spaced	40	Dist.

Randomization of program trials

Massed condition: 2, 1, 3, 4
Spaced condition: 4, 1, 2, 3
Distributed condition: 4, 3, 1, 2

*Reliability days.

Figure 2. Sample schedule sheet showing the condition for each dat of training and the order of training trials for each condition.

Data Sheet

Student No. 13

Name Jim

Teacher Tom

Condition Massed

Date 11/3/80

Day 10

Time start 10:05 stop 10:30

Trial No.	Task	Score	Trial No.	Task	Score
1	Table Set	1	21	Folding	2
2	Table Set	1	22	Folding	2
3	Table Set	1	23	Folding	1
4	Table Set	1	24	Folding	2
5	Table Set	2	25	Folding	1
6	Table Set	1	26	Folding	2
7	Table Set	1	27	Folding	2
8	Table Set	1	28	Folding	1
9	Table Set	1	29	Folding	1
10	Table Set	2	30	Folding	1
11	Pict/Obj. Car	1	31	Comm. Thirsty	1
12	Pict/Obj. Brush	2	32	Comm. Thirsty	1
13	Pict/Obj. Car	0	33	Comm. Thirsty	2
14	Pict/Obj. Brush	2	34	Comm. Thirsty	1
15	Pict/Obj. Car	1	35	Comm. Thirsty	1
16	Pict/Obj. Brush	2	36	Comm. Thirsty	1
17	Pict/Obj. Brush	1	37	Comm. Thirsty	1
18	Pict/Obj. Car	0	38	Comm. Thirsty	1
19	Pict/Obj. Car	0	39	Comm. Thirsty	1
20	Pict/Obj. Brush	1	40	Comm. Thirsty	2

Figure 3. Sample data sheet for the massed condition.

student's responses on a separate data sheet. At the end of the session reliability was computed and expressed as percent by the following formula: number of agreements/number of agreements plus number of disagreements X 100 for all trials in the session.

Data Analysis

The Friedman Two Tailed Test (Conover, 1971), a nonparametric ANOVA, was applied to the means, slopes, and correlations from the correct, incorrect, and refusal scores during each condition (massed, spaced, distributed) for each program for-each student. The statistics utilized during the Friedman Test were derived from the daily data recorded during each training session. The mean and correlation data were computed using the percent responding and the number of training days. The equation for mean calculation was the sum of percent scores per session divided by the total number of sessions. The correlations were computed using the Pearson Correlation Coefficient (Conover, 1971) with y = percent score per session and x = session number. The slopes were computed using the percent responding as the Y coordinate and the actual session number as the X coordinate. These statistics were computed for the correct, incorrect, and refusal scores across the experimental conditions of massed, spaced, or distributed trial sequencing for each phase of all four training programs for each student. This resulted in sixty-four cases for the overall analysis. The analyses were made utilizing the Honeywell 60/66 and the SPSS Program for the Friedman Test (Klecka, Nie, & Hull, 1975).

RESULTS

Reliability

Table 2 shows the reliability measures for each program for each student. The mean reliabilities are displayed by condition: massed,

Table 2

Mean Reliability by Program and Training Condition for All Students

Student Number	Program Number	Mean Reliability				
		Condition				
		Massed	Spaced	Distributed	Program	Overall
1	1	90	100	100	96.7	97.5
	2	100	100	90	96.7	
	3	100	90	100	96.7	
	4	100	100	100	100	
2	1	100	100	100	100	98.35
	2	100	100	100	100	
	3	90	100	100	96.7	
	4	100	100	90	96.7	
3	1	100	100	100	100	97.5
	2	90	100	100	96.7	
	3	100	90	100	96.7	
	4	100	100	90	96.7	
4	1	100	100	100	100	96.7
	2	90	90	100	93.3	
	3	80	100	100	93.3	
	4	100	100	100	100	
5	1	100	100	90	96.7	98.75
	2	100	100	100	100	
	3	100	100	100	100	
	4	95	100	100	98.3	
6	1	90	100	80	90	91.7
	2	90	100	100	96.7	
	3	90	100	80	90	
	4	100	100	70	90	
7	1	100	100	100	100	99.2
	2	100	100	100	100	
	3	100	90	100	96.7	
	4	100	100	100	100	
8	1	100	100	95	98.3	96.7
	2	90	90	85	88.3	
	3	70	100	100	100	
	4	100	100	100	100	

Table 2 (cont.)

Mean Reliability by Program and Training Condition for All Students

Student Number	Program Number	Mean Reliability					
		Condition					
		Massed	Spaced	Distributed	Program	Overall	
9							
	1	100	100	90	96.6		
	2	100	90	100	96.9	93.3	
	3	90	90	90	83.3		
	4	90	100	90	96.6		
10	1	100	100	100	100		
	2	100	70	90	86.6	94.9	
	3	90	100	100	96.6		
	4	90	100	100	96.6		
11	1	100	90	100	96.6		
	2	100	100	100	100	99.15	96.7
	3	100	100	100	100		
	4	100	100	100	100		

spaced, or distributed trial training, by program, and by students. Across all students the reliability scores per conditon ranged from 70% to 100% with 96% of the scores falling between 90% and 100%. Mean reliability scores by program ranged from 83.3% to 100% with 93% falling between 90% and 100%. The mean reliability scores by student ranged from 91.7% to 99.2% with an overall reliability measure of 96.7%.

Insert Table 2 about here

Nonparametric Analysis of Variance

Table 3 shows the results of the application of the Friedman Test (Conover, 1971) to four questions of primary interest. Students' performance of correct, error, and refusal responses were compared across each condition of training. The statistics utilized to describe the students' performance were mean level of responding, slope of responding, and the correlation of the responses within conditions. Three comparisons were significantly different at the .05 level and one comparison approached significance at the .09 level. When comparing only massed vs. distributed trial training for all cases there was a significant difference in mean levels of responding for correct responses and refusals to respond. When massed vs. distributed trial training was compared for motor programs the mean level of correct responding approached significance at the .09 level. When massed vs. distributed trial training was compared for cognitive programs, the mean level of refusals to respond was found to be significantly different. An analysis of the actual scores showed that all of the significant findings were in favor of distributed trials being more effective than massed trial training, i.e.

correct responding was higher and refusals to respond were lower for distributed than for massed trial sequencing.

Insert Table 3 about here

Table 4 shows the results of the application of the Friedman Test to five questions of secondary interest. Students' performance on correct, error, and refusals to respond were compared across each condition of training. The statistics used to describe the responses were the mean level of responding and the slope of the responses. Two comparisons were significantly different at the .025 level and two comparisons approached significance at the .05 level. The .025 level of significance was used given the "post hoc" nature of the analysis. When massed vs. distributed trial training was compared for programs accelerating, the difference in mean levels of responding for correct and refusal responses approached significance at the .05 level. When massed vs. distributed trial training was compared for students living in the home, the mean levels of responding for correct and refusals to respond were significantly different at the .025 level. When the same comparison was made for students living in institutions, no differences were detected. The results found to be significant were all in favor of distributed trial training.

Insert Table 4 about here

Table 3

Results of Nonparametric Analysis of Variance (The Friedman Test) for Comparison of Primary Interest

Comparisons	Cases	df	Response	Statistics	α
Massed vs. Spaced vs. Distributed Trial Training for All Cases	69	2	correct	mean	.15
				slope	.75
				correlation	.95
			error	mean	.35
				slope	.99
				correlation	.50
			refuse	mean	.23
				slope	.80
				correlation	.89
Massed vs., Distributed Trial Training for All Cases	69	1	correct	mean	.04*
				slope	.71
				correlation	.80
			error	mean	.22
				slope	.71
				correlation	.99
			refuse	mean	.05*
				slope	.90
				correlation	.81
Massed vs. Distributed Trial Training for Motor Programs	37	1	correct	mean	.09
				slope	.45
				correlation	.70
			error	mean	.57
				slope	.99
				correlation	.34
			refuse	mean	.70
				slope	.85
				correlation	.99

Comparisons	Cases	df	Response	Statistics	α
Massed vs. Distributed Trial Training for Cognitive Programs	28	1	correct	mean	.51
				slope	.86
				correlation	.51
			error	mean	.25
				slope	.86
				correlation	.86
			refuse	mean	.05*
				slope	.74
				correlation	.51

*Significant at the .05 level in favor or Distributed Trial Training

Table 4

Results of the Nonparametric Analysis of Variance (The Friedman Test) for Comparisons of Secondary Interest

Comparisons	Cases	df	Response	Statistic	α
Massed vs. Distributed Trial Training for Programs Accelerating	34	1	correct	mean	.04*
				slope	.71
			error	mean	.23
				slope	.71
			refuse	mean	.05*
				slope	.90
Massed vs. Distributed Trial Training for Motor Programs Accelerating	9	1	correct	mean	.31
				slope	.73
			error	mean	.5
				slope	.31
			refuse	mean	.73
				slope	.73
Massed vs. Distributed Trial Training for Cognitive Programs Accelerating	23	1	correct	mean	.83
				slope	.83
			error	mean	.83
				slope	.53
			refuse	mean	.40
					.67
Massed vs. Distributed Trial Training for Students Living in The Home	32	1	correct	mean	.01**
				slope	.99
			error	mean	.11
				slope	.48
			refuse	mean	.025**
				slope	.21

Comparisons	Cases	df	Response	Statistic	
Massed vs. Distributed Trial Training for Students Living in Institutions	37	1	correct	mean	.62
				slope	.62
			error	mean	.86
				slope	.86
			refuse	mean	.51
				slope	.18

**Significant at the .025 level in favor of Distributed Trial Training

DISCUSSION

Mean level of responding was the only variable effected by the conditions of massed and distributed trials sequencing. Some guarded conclusions, however, can be made about the comparative effects of massed and distributed trial scheduling on the performance of severely handicapped students.

Related Research

The finding of no difference between distributed and massed trial sequencing for cognitive programs is not supported by the majority of literature utilizing cognitive tasks in the investigation of the distributed trials effect. The tasks used to demonstrate this effect were language skills involving nonfunctional activities such as paired associate responding, free recall, serial recall, and rote learning. The cognitive activities utilized in this study did not include tasks similar to those used in the above studies. Typically the activities included sorting, match to sample, or picture object matching tasks. Some included total cycle tasks such as table setting or turning on a tape recorder. The discrepancies in the types of cognitive tasks utilized in this study make a direct comparison with the trends found in the literature difficult.

The similarity of tasks and the meaning of items have been utilized to investigate the effects of distributed ~~and~~ massed practice. Underwood and colleagues (Underwood, 1953; Underwood & Goad, 1951; Underwood, Kapelak, & Malmi, 1976) demonstrated that college students' learning of nonsense syllables was significantly superior under a distributed condition versus a massed condition as a function of intralist similarity. The more difficult the list, the more the distributed practice condition facilitated learning. In a previous study, Underwood (1952) demonstrated that learning was faster in the distributed conditon for acqusition of a serial list of adjectives with high interlist similarity as compared to a list of low similarity. Similarly, Tsao (1948) demonstrated the differential effect of learning known and unknown items on the distributed versus massed practice effect. College students learned low meaning nonsense syllables significantly better in a spaced rather than a massed condition. Learning of high meaning nonsense syllables, however, was not differentially effected by the massed or spaced conditon. One of the requirements of current program development for the severely handicapped is that the task be functional and age appropriate for the individual student. The teacher develops programs which are useful to the student, utilizing familiar materials and environments. In this respect, the activities and materials employed in this study could be considered to be meaningful or familiar to the learner. If this can be inferred, then the finding of no significant difference between correct performance in distributed and massed trials sequencing is consistent with previous findings.

This study found no difference in error responding for massed and distributed trials in any of the analyses. In his studies with college students, Underwood (1953) found that the rate of overt error in the learning of nonsense syllable was inversely related to the length of the intertrial interval. A comparison of massed and distributed practice would indicate that distributed practice resulted in fewer errors than did massed practice. In an earlier study Underwood (1952) noted that college students made more errors with tasks requiring continuous performance, such as color naming, then with tasks requiring intermittent responding, such as intermittent symbol cancellation. In a study with handicapped individuals, Dent and Johnson (1964) found error responding to be reduced by half for performance under distributed conditions versus massed conditions. A possible explanation for the finding of no significance difference for error responding between the massed and distributed conditions in this study is the inclusion of the measure of no responding or refusal to respond. Recently Dunlap and Koegel (1980) have demonstrated that autistic children show an increase in "no responses" across a constant task presentation of typical cognitive programs. A varied task presentation produced a decrease or removal of "no responses". Constant task presentation was defined as repeated presentations of a single experimental task. Varied task presentation was such that a task was never repeated more than two times in a row and was interspersed or distributed among a variety of other tasks. Dunlap and Koegel discussed these findings in terms of the process of varying the tasks affecting the student's motivation to respond rather than the student's ability to learn the task. This study provides support for the results that mean levels of responding, rather than the slope of the response, were significantly different for all cases when comparing distributed and massed trial sequencing.

Previous studies dealing with the acquisition and retention of motor skills repeatedly demonstrated the superiority of distributed trials sequencing over massed trials sequencing. Typically the studies utilized college students on nonfunctional tasks and measured performance in terms of number of trials to criterion or number correct per session. Chasey (1976), while working with retarded boys, demonstrated that not only did those students learning the fine motor task under distributed practice require fewer trials to criterion, but that the group experiencing the most distribution (30 seconds work and two minutes rest) had significantly better retention than the groups experiencing massed trials or trials with shorter distribution of trials. An analysis of motor programs in this study did not reveal any difference in performance of correct responding for the distributed and massed conditions.

The type of tasks typically used have included a motor pursuit task (Dore & Hilgard, 1938; Cook & Hilgard, 1949, Reynolds & Adams, 1953; and Denny, Frisbey, & Weaver, 1955), a stabilometer task (Chasey, 1976), printing the alphabet up-side-down (Kientzle, 1946; Kimble, 1949a, 1949b), dot connecting (Underwood, 1961) running a maze and a puzzle box (Erickson, 1942), and a foot tracking task (Whitley, 1970). All the tasks, except running the maze and the puzzle box, defined a trial as a period of time rather than a discrete trial. For example, performance on the stabilometer was broken into trials of thirty seconds of work and then differing periods of rest. In this present study typical motor tasks included cleaning a surface area, holding a pencil correctly, folding towels, or putting down a cup without spilling. A discrete trial presentation procedure was used in which a trial was define as the complete cycle from the presentation of a stimulus to the response, correct procedure (if needed), and reinforcement. The student was never in a situation requiring continuous work on a task without the interruption of an eliciting cue and some type of structured consequence. The difference in trial structure between this and previous studies makes direct comparisons difficult.

None of the studies dealing with the comparison of massed versus distributed trial scheduling addressed the question of living situation or daily environment encountered by the student. This study found performance for correct and refusal responding to be significantly different for individuals living in the home. A related study (Dunlap and Koegel, 1980) demonstrated that autistic children increase correct responding and decrease refusal responding under a varied task condition. These children were reported to have been living at home and attending public school at the time of the study. This is the same condition experienced by the set of students identified as "living in the home" involved in the present study. Even though Dunlap and Koegel did not discuss the question of living situation, they did suggest that the key factor to the differentation in performance between the constant and varied task conditon was the element of variety. It could be hypothesized that children living in the home encounter, become familiar with, and respond more positively to variation in schedules that do children living in institutions.

Implications

A distributed trials sequencing strategy can be added to the technology of instruction for severely handicapped students as a result of the findings in this study. Although the acquisition slopes showed no differences across the three training strategies, distributed trial sequencing did generally product higher mean levels of responding.

It should be noted that no attempt is being made to substitute massed trials sequencing with a distributed trial strategy. The experimental literature (Dore & Hilgard, 1938; Ericksen, 1942; Kimble, 1949a; Reynolds & Adams, 1953) indicates some combination of the two procedures may produce the best learning as well as retention. Currently, the general assumption is that massed trials training would be most efficient during the initial phases for rapid acquisition. Generalization and the functional use of the skill would probably be learned most efficiently during the distributed trials sequencing of functional skills.

The present study focused on the question of the strategy of trial sequencing. Application is directed to the practice of teaching skills in settings and at the time when the performance of a particular skill naturally occurs. This involves separating training trials across the day, the strategy of distributed trial sequencing, and is a departure from the traditional instructional approach of twenty minute training sessions with 10 to 15 trials per session on an isolated skill. Anecdotal reports from the teachers utilizing distributed trial sequencing indicate a positive response to this procedure. The implementation is described as more enjoyable for the teacher as well as the student than massed trial sequencing. In addition, distributed trial sequencing seems to allow for greater integration of skills through the day so the functional relationship between the skills can be taught as well as the individual skills.

Reference Notes

1. Holvoet, J., & Utley, B. Implementing a functional curriculum with severely handicapped. Paper presented at Kansas Statewide Inservice Conference, Topeka, Kansas, 1978.

2. Gargagliano, C. The effects of massed versus distributed practice on the acquisition and maintenance of basic sight words. Working Paper Number 5, Experimental Education Unit, Child Development and Retardation Center, University of Washington, Seattle, Washington, 1974.

3. Helmstetter, E. *The effects of distributed practice on learning by a severely handicapped adolescent*. Unpublished master's thesis, The University of Kansas, 1980.

References

Adams, J.A. Warm-up decrement in performance on the pursuit rotor. American Journal of Psychology, 1952, 55, 404-414.

Baumeister, A.A., & Berry, M. Distribution of practice and specificity of learning in normals and retardates. American Journal of Mental Deficiency, 1976, 72, 227-231.

Barlow, D. H. & Hayes, S.C. Alternating treatments design: One strategy for comparing the effects of two treatments in a single subject. Journal of Applied Behavior Analysis, 1979, 12, 199-210.

Brown, L., Nietupski, J., & Hamre-Nietupski, S. Criterion of ultimate functioning. In M.A. Thomas (Ed.) Hey, don't forget about me. The Council for Exceptional Children, Reston: VA, 1976.

Carron, A.V. Performance and learning in a discrete motor task under massed vs. distributed practice. The Research Quarterly, 1969, 40, 481-489.

Chasey, W.C. Distribution of practice, learning and retention. Perceptual and Motor Skills, 1976, 43, 159-164.

Conover, W.J. Practical nonparametric statistics. New York: John Wiley & Sons, Inc., 1971.

Cook, B.S., & Hilgard, E.R. Distributed practice in motor learning: Progressively increasing and decreasing rests. Journal of Experimental Psychology, 1949, 39, 196-172.

Deese, J. The psychology of learning. New York: McGraw-Hill Book Company, 1958.

Denny, M.R., Frisbey, N., & Weaver, Jr., J. Rotary pursuit performance under alternate conditions of distributed and massed practice. Journal of Experimental Psychology, 1955, 49, 48-54.

Dent, H., & Johnson, R. The effects of massed versus distributed practice on the learning of organic familial defectives. American Journal of Mental Deficiency, 1964, 68, 533-536.

Dore, L.R. & Hilgard, E.R. Spaced practice as a test of Snoddy's two processes in mental growth. Journal of Experimental Psychology, 1938, 23, 359-374.

Duncan, C.P. The effect of unequal amounts of practice on motor leariniing before and after rest. Journal of Experimental Psychology, 1951, 42, 257-264.

Dunlap, G. & Koegel, R.L. Motivating autistic children through stimulus variation. Journal of Applied Behavior Analysis, 1980, 13, 619-627.

Elmes, D.G., Sanders, L.W., & Dovel, J.C. Isolation of massed and distributed items. Memory and Cognitiion, 1973, 1, 77-79.

Ericksen, S.C. Variability of attack in massed and distributed practice. Journal of Experimental Psychology, 1942, 31, 339-345.

Guess, D., Horner, R., Utley, B., Holvoet, J., Maxon, D., Tucker, D., & Warren, S. A functional curriculum sequencing model for teaching the severely handicapped. AAESPH Review, 1978, 3, 202-215.

Holvoet, J., Guess, D., Mulligan, M., & Brown, F. The individualized curriculum sequencing model (II): A teaching strategy for severely handicapped students. Journal of the Association for the Severely Handicapped, 1980, 5, 337-352.

Hovland, I.C. Experimental studies in rote-learning theory, V. Comparison of distribution of practice in serial and paired associate learning. Journal of Experimental Psychology, 1939, 26, 622-633.

Hovland, I.C. Experimental studies in rote-learning theory, VII. Distribution of practice with varying lengths of lists. Journal of Experimental Psychology, 1940a, 27, 271-284.

Hovland, I.C. Experimental studies in rote-learning theory, VI. Comparisons of retention following learning to same criterion by massed and distributed practice. Journal of Experimental Psychology, 1940b, 27, 568-587.

Kientzle, M.J. Properties of learning curves under varied distributions of practice. Journal of Experimental Psychology, 1946, 36, 187-211.

Kimble, G.A. Performance and reminiscence in motor learning as a function of the degrees of distribution of practice. Journal of Experimental Psychology, 1949a, 39, 500-510.

Kimble, G.A An experimental test of a two-factor theory of inhibition. Journal of Experimental Psychology, 1949b, 39, 15-23.

Kimble, G.A., & Bilodeau, E.A. Work and rest as variables in cyclical motor learning. Journal of Experimental Psychology, 1949, 39, 150-157.

Klecka, W.R., Nie, N.H., & Hull, C.H. Statistical packages for the social sciences. Primer. McGraw Hill, Inc., 1975.

Madsen, M.C. Distribution of practice and level of intelligence. Psychological Reports, 1963, 13, 39-42.

Mulligan, M., Guess, D., Holvoet, J., & Brown, F. The individualized curriculum sequencing model (I): Implications from research on massed, distributed, and spaced trial training. Journal of the Association for the Severely Handicapped, 1980, 5, 325-336.

Neef, N.A., Iwata, B.A., & Page, T.J. The effects of known item interspersal on acquisition and retention of spelling and sight reading words. *Journal of Applied Behavior Analysis*, 1977, 10, 738.

Reynolds, B., & Adams, J.A. Effect of distribution and shift in distribution within a single training session. *Journal of Experimental Psychology*, 1953, 46, 137-145.

Stelmach, G.E. Efficiency of motor learning as a function of intertrial rest. *The Research Quarterly*, 1969, 40, 198-202.

Tsao, J.C. Studies in spaced and massed learning, II: Meaningfulness of material and distribution of prctice. *Quarterly Journal of Experimental Psychology*, 1948, 40, 198-202.

Underwood, B.J. Studies in distributed practice: IV. The influence of rest interval activity in serial learning. *Journal of Experimental Psychology*, 1952, 43, 329-341.

Underwood, B.J. Studies in distributed practice: VIII. Learning and retention as a function of interlist similarity. *Journal of Experimental Psychology*, 1953, 45, 133-141.

Underwood, B.J. Distributed practice on the Tsai-Partington Numbers Test. *Perceptual and Motor Skills*, 1961, 12, 325-326.

Underwood, B.J. & Goad, D. Studies of distributed practice: I; The influence of intra-list similarity in serial learning. *Journal of Experimental Psychology*, 42, 1951, 125-134.

Underwood, B.J., Kapelak, S.M. & Malmi, R.A. The spacing effect: Additions to the theoretical and empirical puzzles. *Memory and Cognition*, 1976, 4, 391-400.

Whitley, J.D. Effects of practice distribution on learning of a fine motor task. *The Research Quarterly*, 1970, 41, 576-582.

Woodworth, R.S. & Schlosberg, H. *Experimental psychology*. New York: Henry Holt and Company, 1954.

The Effects of Number and Composition of Interspersed Trials on the Learning of Severely Handicapped Students[1,2]

Jennifer Holvoet, Ph.D.
University of Kansas
Lawrence, Kansas

Lesley Fernandez, M.S. ED.
Kansas Neurological Institute
Topeka, Kansas

1982

---

[1] The authors wish to thank the SMH students, teachers and administrators of USD 501, Topeka for their support and cooperation during this study.

[2] Portions of this paper were prepared as part of the work pursuant to the "Kansas Individualized Curriculum Sequencing Project" funded as a Handicapped Children's Model Project, U.S. Office of Education (G00-800-1721) from 9-1-80 to 9-1-83.

Over the past 15 years, the focus of curriculum development for the severely handicapped population has been on identifying relevant skills to teach, specifying and breaking down these skills so that the results of teaching can be quantified, and the development of strategies which could be coupled with the teaching process (e.g., consistent cues, prompting strategies, positioning). The equally important issue of practice and how practice of different skills should be temporally arranged, however, is only now beginning to be addressed. It appears that in the absence of information about this variable, and in response to the call for data-based instruction and the adoption of techniques from experimental literature, most teachers of severely handicapped students have used a massed practice approach to scheduling training trials. A massed trial practice schedule is defined as one in which the training trials of a single instructional program occur so closely together in time that no other behavior is likely to occur between them (Mulligan, in press). For example, it is not uncommon to see a teacher giving a student 10-20 training trials of a communication program from 10:00-10:15 each day with less than two seconds between trials.

However, some investigators (Dunlap & Koegel, 1980; Helmstetter, Note 1; Mulligan, in press; Neef, Iwata, & Page, 1977) have found that the use of interspersed practice facilitates learning. An interspersed practice schedule is one where the training trials of a single instructional program are separated by a trial or trials of one or more other programs. The interspersed programs (i.e., the ones that are taught between the trials of the target task) can be either already learned skills or new skills. Interspersed practice is synonymous with distributed practice as defined by Mulligan (in press).

Using moderately mentally retarded subjects, Neef, Iwata and Page (1977) studied the effects of interspersal training where ten previously learned words were alternately presented with each of ten test words. These investigators found that both acquisition and retention were facilitated by the interspersal procedures. Further investigation by these authors (1980) compared interspersal trial training with heavily reinforced massed trial training and found that the interspersed practice resulted in superior acquisition as well as superior short- and long-term retention. In addition, the students expressed a preference for the interspersed condition in spite of the difference in tangible reinforcement.

In both the previous studies, the interspersed items had already been learned by the subjects. This, however, is not the way most curricula are arranged. Generally, students are expected to learn several new tasks concurrently. Helmstetter (Note 1), therefore, studied the effect of training a novel task during the intertrial interval of three tasks taught to a deaf-blind severely retarded child. The novel task was alternated with trials of each of the other tasks during interspersal training. The data indicated that interspersed practice was superior to massed practice in terms of acquisition on two of the tasks and was equivalent to massed practice on the other task.

Mulligan (in press) evaluated the effects of massed, distributed and spaced practice on the acquisition of cognitive and motor skills in 11 severely retarded students. The distributed condition consisted of interspersing a trial from each of three novel tasks in the intertrial interval. Using a two-tailed test, she found that mean levels of responding in the distributed and massed conditions were significantly different for correct and refusal responses. There were less refusal responses and more correct responses in the distributed condition. This distribution effect was shown to be more pronounced for students who lived at home than for those who lived in an institution. She found no differences between the spaced, distributed, or massed conditions in how quickly the student's learned the tasks.

Dunlap and Koegel (1980) compared interspersed and massed trial methods of teaching autistic children a task. In this study, other unlearned tasks from the child's curriculum were used for interspersal. They found that the massed condition resulted in a decline in correct responses on the target task, whereas improved and stable responding occurred with the interspersal method.

It therefore appears that the literature supports a curriculum approach for handicapped persons where several instructional programs are taught using interspersed practice. This study looks at some of the parameters of the interspersal technique as it is used with severely handicapped students. Specifically, it is an investigation of whether the number of interspersed tasks affects learning; and whether the effect of interspersed practice is different when single trials of several tasks are interspersed instead of interspersing multiple trials of the same task.

Two experiments were conducted. The essential difference between the two experiments was that the first evaluated the effect of different numbers of interspersed trials when these trials were drawn from several different tasks, while the second experiment evaluated the effect of interspersing different numbers of trials of a single task.

EXPERIMENT 1

Experiment 1 was conducted to compare learning under three conditions. These conditions were: 1) no interspersed trials; 2) interspersing a trial from each of two different tasks; and 3) interspersing a trial from each of four different tasks.

Method

Subjects and Settings

Subject 1 (John) was a 10-year-old male with limited vision and hearing, moderate quadriplegic hypotonia, no speech, and frequent minor and major epileptic seizures. He performed at the 0-18 month level on tests of adaptive behavior such as the Callier-Azusa Scale (Stillman, 1978) and was diagnosed as profoundly retarded. Subject 2 (Jane) was a 15-year-old female with no speech and visual problems which were corrected by glasses. She was classified as severely retarded, scoring in the 40th percentile on the TARC Assessment Inventory for Severely Handicapped Children (Sailor & Mix, 1975). Both subjects resided in an institution for the mentally retarded and attended different self-contained public school classrooms for the severely multiply handicapped. The study was conducted at the subject's desk within each subject's classroom during school hours. Other students and teachrs within the classroom continued with their daily routines while the study was in progres.

Training Tasks

The training tasks were divided into two categories for purposes of determining the effects of the interspersal procedure. The tasks presented prior to the intertrial interval were called TARGET TASKS, and the tasks presented during the intertrial interval were called INTERSPERSED TASKS.

Tasks were drawn from several instructional domains such as communication, fine motor skills, self-help, etc.

Two target tasks and eight interspersed tasks were identified for each subject. These tasks were then divided into two training sets. A training set consisted of a target task and four interspersed tasks. Each subject thus had two sets of training tasks, designated Set A and Set B. The tasks used with each subject and the designation of these as target/interspersed are outlined in Table 1. Complete instructional programs were written for each of these tasks, either by the classroom teacher or the experimenter.

Insert Table 1 about here

For both target and interspersal tasks, the students were given social praise, pats, and a token or edible for every correct response whether it was independently initiated or prompted. The students were, however, given only social praise if physical guidance was needed to obtain a correct response.

Experimental Procedures

To determine entry levels, each student was pretested on all of the target and interspersed programs in both training sets before the experimental conditions were begun. The pretests were done using a massed trial format and were conducted over two days. Each experimental session was composed of 10 trial sets. A trial set consisted of the presentation of one target task trial and a four-minute intertrial interval (with or without the interspersed tasks presented during this interval). Descriptions of a trial-set for each of the three conditions used in this experiment follow:

Table 1

Training Tasks Used in Experiment One

STUDENT	TRAINING SET DESIGNATION	TARGET TASK	INTERSPERSED TASK
Jane	Set A	Number sequencing	1. Button shirt 2. Cut food 3. Set 1 place setting 4. Demonstrate knowlege of prepositions (in front of, in back of, beside)
Jane	Set B	Trace name	1. Sign "cracker/book" 2. Collate 4 page manual 3. Demonstrate knowledge of object position (sideways, upside down, right side up) 4. Use of spray bottle
John	Set A	Shirt off	1. Discrimination of yellow 2. Hand-to-hand transfer 3. Sign "glass" 4. Take off lid
John	Set B	Sorting	1. Put cup down without spilling 2. Body part identification (hand) 3. Use of string to obtain object 4. Using one finger to point

Condition 1 - No interspersed tasks (Spaced): One trial of the target task (e.g., Trace name) and a 4-minute rest during the intertrial interval;

Condition 2 - Two interspersed tasks (D-2): One trial of the target task (e.g., Trace name) and a 4-minute intertrial interval during which one trial of each of two different interspersal tasks (e.g., Signing and Collating) were presented; and

Condition 3 - Four interspersed tasks (D-4): One trial of the target task (e.g., Trace name) and a 4-minute intertrial interval during which one trial of each of four different interspersal tasks (e.g., Signing, Collating, Object position, and Use of spray bottle) were presented.

The way the task-sets were organized into a session is illustrated in Figure 1.

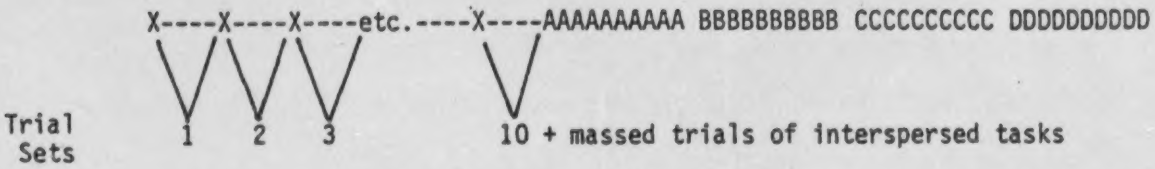
Insert Figure 1 about here

As can be seen in Figure 1, in the Spaced and D-2 conditions, where all or some of the interspersal tasks were not presented during the intertrial interval, 10 trials of each of these tasks were conducted in a massed trial format after the completion of the trial sets. This made the amount of practice on all tasks the same across conditions.

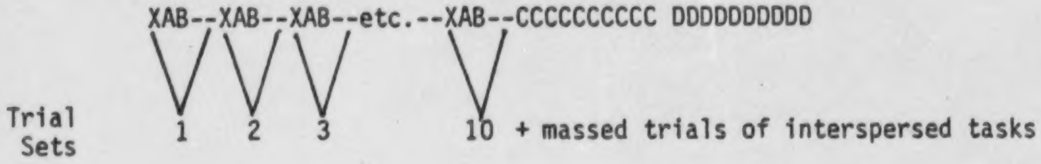
In the D-2 and D-4 conditions, if the 4-minute intertrial interval had not elapsed by the time the student had completed the trials of the interspersed tasks, the student was allowed to rest until the end of the interval. In general, there was a 2-minute rest in the D-2 condition and no rest in the D-4 condition. The presentation and completion of the interspersed tasks never exceeded the 4-minute intertrial interval.

Figure 1: Examples of a training session for each experimental condition where X indicates a trial of the target task , the dashed lines indicate waiting time during the intertrial interval, and A, B, C, and D indicate the four different interspersed tasks.

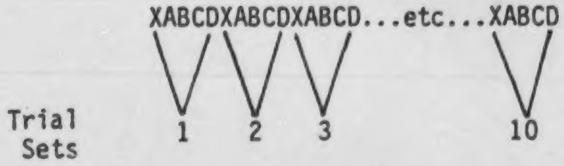
Condition 1: Spaced Practice


X----X----X----etc.----X----AAAAAAAAAA BBBBBBBBBB CCCCCCCCCC DDDDDDDDDD
Trial Sets
1
2
3
10 + massed trials of interspersed tasks

Condition 2: Two Interspersed Tasks (D-2)


XAB--XAB--XAB--etc.--XAB--CCCCCCCCCC DDDDDDDDDD
Trial Sets
1
2
3
10 + massed trials of interspersed tasks

Condition 3: Four Interspersed Tasks (D-4)



Experimental Design

The effects of the different number of interspersed tasks on learning were evaluated using a multiple schedule design. Hersen and Barlow (1978) note that this design involves differentially varying stimulus conditions on the same behavior. This was accomplished in this study by varying the number of interspersed tasks across sessions. The alternation of the experimental conditions were predetermined in that all permutations of the three conditions were used, and no more than two sessions of the same condition occurred sequentially. The schedule of experimental conditioned differed for each subject and each training set. Table 2 shows these schedules of experimental conditions.

Insert Table 2 about here

Reliability

Interobserver reliability was assessed at least once in every condition for each subject. A second observer scored independently and simultaneously with the experimenter. Reliability was calculated by dividing the number of trials in which the two observers agreed by the total number of trials observed during the session and multiplying by 100.

Results

Reliability

Jane's mean reliability scores, calculated across programs, were 99%, 100%, and 97% respectively for the Spaced, D-2 and D-4 conditions. The range of reliability for Jane's data was 90-100%.

Table 2

Schedule of Experimental Conditions for Experiment 1

SUBJECT	TRAINING SET DESIGNATION	SESSION NUMBER	EXPERIMENTAL CONDITION
John	Set A	1	Spaced
		2	D-2
		3	D-4
		4	D-2
		5	Spaced
		6	D-4
		7	D-2
		8	D-4
		9	Spaced
		10	Spaced
		11	D-4
		12	D-2
John	Set B	1	D-4
		2	D-2
		3	Spaced
		4	Spaced
		5	D-4
		6	D-2
Jane	Set A	1	Spaced
		2	D-2
		3	D-4
		4	D-2
		5	Spaced
		6	D-4
		7	D-4
		8	D-2
		9	Spaced
		10	D-2
		11	D-4
		12	Spaced
		13	Spaced
		14	D-2
		15	D-4
		16	Spaced
		17	D-2
		18	D-4
		19	D-4
		20	Spaced
		21	D-2

(Table continued on next page)

Table 2 (continued)

SUBJECT	TRAINING SET DESINGATION	SESSION NUMBER	EXPERIMENTAL CONDITION
Jane	Set B	1	D-2
		2	D-4
		3	Spaced
		4	D-2
		5	D-4
		6	Spaced
		7	D-2
		8	Spaced
		9	D-4

The mean reliability scores for John, calculated across programs, were 98%, 96%, and 96% for the Spaced, D-2 and D-4 conditions respectively. The range of reliability for these data was 80-100%.

The Effects of Interspersing Zero, Two or Four Tasks

The effect of interspersing zero, two or four tasks was first measured by comparing data obtained under the three experimental conditions from the TARGET tasks. As can be seen in Table 3, two of the target tasks were learned best under the D-4 condition, one was learned best under the D-2 condition and one was learned best under the Spaced condition. There did not appear to be a consistent effect across either subject.

Insert Table 3 about here

The effect of interspersing two or four tasks could also be measured by comparing the acqusition data from the INTERSPERSED tasks under the conditions D-2 and D-4. This is possible because in the D-2 condition each interspersed task trial was followed by a trial of another interspersed task and a trial of the target task (though not necessarily in that order). For example, referring back to Figure 1, it can be seen that interspersed task B is followed by a rest, then a trial of target task X, then a trial of interspersed task A. In the D-4 condition, each interspersed task trial was followed by a trial of each of three other interspersed tasks and a trial of the target task (not necessarily in that order). Again referring back to Figure 1, one can see that interspersed task B is followed by trials of interspersed tasks C and D, then a rest, then a trial of target task X, then a trial of interspersed task A. The data in Table 3 show that the D-4 condition resulted in slightly higher scores for six of the eight interspersed tasks.

Table 3

The Effect of Interspersing Zero, Two or Four Different Tasks

STUDENT	TYPE OF TASK	MEAN PERCENT CORRECT		
		SPACED	D-2	D-4
Jane	Target	76%	83%	76%
	Target	73%	80%	90%
	Interspersed	--	67%	77%
	Interspersed	--	63%	70%
	Interspersed	--	81%	84%
	Interspersed	--	86%	89%
John	Target	85%	75%	75%
	Target	68%	56%	70%
	Interspersed	--	68%	43%
	Interspersed	--	33%	38%
	Interspersed*	--	70%	60%
	Interspersed*	--	45%	50%

* These data based on only two data points per condition

Looking at the data across both TARGET and INTERSPERSED tasks, it can be seen that the highest performance was achieved under the D-4 condition on 8 of the 12 tasks, under the D-2 condition on three tasks, and one task was learned best under the spaced condition. In particular, Jane's performance was enhanced by the D-4 condition, learning five of her six tasks better under that condition.

EXPERIMENT 2

Experiment 2 was conducted to compare the effects of interspersing zero, two or four trials of the same task between trials of a target task. It differs from Experiment 1 where trials from several different tasks were interspersed.

Method

Subjects and Settings

The subjects and settings used in Experiment 2 are the same as those used in Experiment 1.

Training Tasks

As in Experiment 1, the training tasks were divided into TARGET TASKS AND INTERSPERSED TASKS which were drawn from several domains. John's target tasks were the same as those used in Experiment 1. Only one of Jane's target tasks from Experiment 1 was used in this experiment due to time limitations. The tasks used with each subject and the designation of these as target or interspersed are outlined in Table 4. Complete instructional programs were written for all tasks, either by the classroom teacher or the experimenter.

Insert Table 4 about here

Reinforcement and program changes were implemented as in Experiment 1.

Experimental Procedures

Each session consisted of 10 trial-sets. Descriptions of the trial-sets for each of the three conditions in this experiment follow:

Condition 1 - No interspersed trials (Spaced): One trial of the target task (e.g., Trace name) and a 4-minute rest during the intertrial interval;

Condition 2 - Two interspersed trials from the same program during the intertrial interval (S-2): One trial of the target task (e.g., Trace name) and a 4-minute intertrial interval in which two trials of a second task (e.g., Classification) were presented;

Condition 3 - Four interspersed trials from the same program during the intertrial interval (S-4): One trial of the target task (e.g., Trace name) and a 4-minute intertrial interval in which four trials of a second task (e.g., Classification) were presented.

Figure 2 illustrates complete session under the three experimental conditions. In the Spaced condition, where the task used for interspersal was not presented during the intertrial interval, 10 trials of this task were conducted in massed trial format after the completion of the trial sets.

Insert Figure 2 about here

In the S-2 and S-4 conditions, if the 4-minute intertrial interval had not elapsed by the time the student completed the trials of the

Table 4

Training Tasks Used in Experiment Two

STUDENT	TRAINING SET DESIGNATION	TARGET TASK	INTERSPERSED TASK
Jane	Set A	Trace name	Classification
John	Set A	Shirt off	Vocalization on command
John	Set B	Sorting	Put object in bucket

Figure 2: Examples of a training session for each experimental condition where X indicates a trial of the target task, the dashed lines indicate waiting time during the intertrial interval, and Y indicates trials of the interspersed task.

Condition 1: Spaced Practice

```
         X----X----X----etc.-----X----YYYYYYYYYY
Trial     \  /  \  /  \  /        \  /
 Sets      1     2     3           10 + massed practice of the
                                        interspersed task
```

Condition 2: Two Interspersed Trials (C-2)

```
         XYY--XYY--XYY--etc.--XYY--
Trial     \ /  \ /  \ /        \ /
 Sets      1    2    3          10
```

Condition 3: Four Interspersed Trials (C-4)

```
         XYYYYXYYYYXYYYY..etc...XYYYY
Trial     \ /  \ /  \ /          \ /
 Sets      1    2    3            10
```

interspersed task, the student was allowed to rest until the end of the interval. In general, there was a 2-minute rest in the S-2 condition and no rest in the S-4 condition. All trials of the interspersed task were always completed within the 4-minute intertrial interval.

Experimental Design

The experimental design used in this experiment was the multiple schedule design. Table 5 shows the schedule of experimental conditions.

Insert Table 5 about here

Reliability

Reliability was computed the same way and by the same formula used in Experiment 1.

Results

Reliability

Mean reliability scores, calculated across Jane's programs, were 95%, 100%, and 100% for the Spaced, C-2, and C-4 conditions respectively. The range of reliability for Jane's data was 90-100%.

John's mean reliability scores, calculated across programs, were 100%, 99%, and 97% respectively for the Spaced, C-2, C-4 conditions. The range of these data was from 90-100%.

The Effect of Zero, Two or Four Interspersed Trials

Table 6 illustrates the effects of interspersing different numbers of trials from a second task on the TARGET tasks. It can be seen that the best performance on all three target tasks was obtained under either the S-2 or the S-4 conditions. John performed best under the S-2 condi-

Table 5

Schedule of Experimental Conditions for Experiment 1

STUDENT	TRAINING SET DESIGNATION	SESSION NUMBER	EXPERIMENTAL CONDITION
John	Set A	1	S-4
		2	Spaced
		3	S-2
		4	S-4
		5	S-2
		6	Spaced
		7	S-2
		8	Spaced
		9	S-4
		10	S-4
		11	S-2
		12	Spaced
John	Set B	1	S-4
		2	S-2
		3	Spaced
		4	S-4
		5	Spaced
		6	S-2
		7	S-2
		8	Spaced
		9	S-4
		10	Spaced
		11	S-4
		12	S-2
		13	Spaced
		14	S-2
		15	S-4
		16	S-2
		17	S-4
		18	Spaced
Jane	Set A	1	S-4
		2	S-2
		3	Spaced
		4	Spaced
		5	S-2
		6	S-4
		7	Spaced
		8	S-4
		9	S-2
		10	S-2
		11	Spaced
		12	S-4
		13	S-2
		14	Spaced
		15	S-4

tion on both target tasks. Jane, however, performed best under the S-4 condition. The effects of interspersal on the INTERSPERSED task could not be determined since there were differing numbers of trials of this task in the different experimental conditions.

Insert Table 6 about here

Discussion

This study was designed to investigate the effect of using different numbers of interspersal items on the learning of severely handicapped students. Skills which the students had not previously learned were used for both the target and the interspersed tasks. The differences in mean acquisition among the conditions generally were quite small, but seemed to follow a consistent pattern. In both experiments, the students showed better learning under at least one of the interspersed conditions than they did under the spaced condition, for the majority (14 out of 15) of their tasks. Jane learned all but one of her tasks better under both interspersed conditions. This occurred regardless of whether the interspersal items were varied, as in Experiment 1, or kept constant, as in Experiment 2. The other task was learned best under the S-4 condition, but the S-2 condition was worse than the spaced condition. Six of her seven tasks were learned best under conditions where four items were interspersed. John learned all but one of his tasks better under an interspersed condition. The other was learned better under the spaced condition in Experiment 1. When the interspersal involved repeated trials of a second program (Experiment 2), John did best in the S-2 condition. It should also be pointed out that both interspersed practice

Table 6

The Effect of Interspersing Zero, Two or Four Trials of the Same Task

STUDENT	TYPE OF TASK	MEAN PERCENT CORRECT		
		SPACED	S-2	S-4
Jane	Target	52%	50%	70%
John	Target	17%	35%	70%
	Target	60%	70%	63%

conditions in Experiment 2 resulted in better performance for John than the spaced condition. This was not true of Experiment 1, where superior performance in one interspersed condition did not necessarily mean that performance in the other interspersed condition would be better than that seen with spaced practice. When the interspersal involved trials of several different programs (Experiment 1), the results were mixed. John learned three of the tasks better under the D-2 condition and three of the tasks better under the D-4 condition.

Overall, the data seem to indicate that the differences between learning under spaced and interspersed practice are small, but consistent. It appears that interspersed practice is usually superior to spaced practice whether the interspersed items are constant or varied. This effect, however, seems to be more clear-cut for the higher functioning student.

The number of interspersed trials seems to be an important variable, but the effect seems to differ with the individual student. The use of four-item interspersal consistently resulted in the best performance of the severely retarded subject. On the other hand, the two-item interspersal procedure seemed to be more effective with the profoundly handicapped student. Obviously, with such a small sample size, it would not be possible to determine whether this is a general phenomenon or not. Nevertheless, these data are in accord with a common-sense analysis. It would seem more sensible to combine fewer tasks when teaching a profoundly handicapped student than might be used when working with a less retarded individual.

These results support the use of interspersed practice in the classroom, and indicate that the teacher can effectively use two or more interspersed trials. These data also seem to indicate that the teacher can choose to use several trials from one program or one trial from each of several different programs as the interspersed items. Further research needs to be done to compare these two strategies (different and constant interspersal) to determine whether either is superior to the other. Further research also needs to be done to ascertain the validity of the finding that the most effective number of interspersal items needs to be individually determined and may be related to overall student functioning.

Reference Notes

1. Helmstetter, E. The effects of distributed practice on learning by a severely handicapped adolescent. Unpublished Master's Thesis. The University of Kansas, 1980.

REFERENCES

Dunlap, G., and Koegel, R.L. Motivating autistic children through stimulus variation. Journal of Applied Behavior Analysis, 1980, 13, 619-627.

Hersen, M., and Barlow, D.H. Single case experimental designs. New York: Pergamon, 1976.

Neef, N.A., Iwata, B.A., and Page, T.J. The effects of known-item interspersal on acquisition and retention of spelling and sight reading words. Journal of Applied Behavior Analysis, 1977, 10, 738.

Neef, N.A., Iwata, B.A., and Page, T.J. The effects of interspersal training versus high density reinforcement on spelling acquisition and retention. Journal of Applied Behavior Analysis, 1980, 13, 153-158.

Sailor, W. and Mix, B.J. TARC Assessment System. Lawrence, KS: H & H Enterprises, 1975.

Stillman, R. (Ed.) The Callier-Azusa Scale. Dallas, TX: Callier Center fpr Communication Disorder, 1978.

Effects of Distributed Practice Instruction on the Cross-Setting Generalization of Severely Handicapped Adolescents

Darrel M. Brewer

Jennifer F. Holvoet

Marilyn Mulligan

Diar Carr

University of Kansas
1983

This paper was prepared as a part of the work pursuant to the "Kansas Individualized Curriculum Sequencing Project" funded as a Handicapped Children's Model Project, U.S. Office of Education (G00-800-1721) from 9-1-80 to 9-1-83.

A major objective of most educational programs for severely handicapped individuals is to provide students with some degree of functional independence (Clark, 1979). It is apparent that if functional independence is to be achieved, the skills the student learns must be exhibited across time, persons, and settings (Stokes & Baer, 1976). When this happens, the student is said to have shown stimulus generalization (Warren, Rogers-Warren, Baer, & Guess, 1980). The training of severely handicapped students is, however, complicated by the fact that, "no inference can be made that training to criterion on any task in one situation will result in criterion performance in similar but different situations requiring similar or slightly different actions" (Brown, Nietupski, & Hamre-Nietupski, 1976, pg. 6). Therefore, it is evident that stimulus generalization should be a programmed element of instructional curricula for severely handicapped students.

A variety of techniques have been employed to promote generalization of skills across settings. Brown et al. (1976) suggested that generalization of a skill could be successfully taught to this population by an instructional strategy that involves training the skill to criterion in one environment, then soliciting (and retraining, if necessary) that skill in several other environments, with different teachers and slightly different materials. This successive training strategy has, in fact, been shown to be successful (Allen, 1973; Griffiths & Craighead, 1972), but involves a tremendous expenditure of time on the part of the student and the teacher. Walker and Buckley (1972) suggested that generalization of skills to nontraining settings may be facilitated by establishing common stimulus elements between training and nontraining settings. From this suggestion has grown a literature stressing the importance of training functional tasks in settings that are as similar as possible to those settings where the behavior will eventually need to be performed (Brown, Branston, Hamre-Nietupski, Pumpian, Certo, & Gruenewald, 1979; Guess, Horner, Utley, Holvoet, Maxon, Tucker, & Warren, 1978; Wilcox & Bellamy, 1982). This stress on functional training has, in fact, reduced some of the generalization problem since the students who are trained to do functional tasks in functional environments may not need to show as extensive generalization as those trained nonfunctionally. Functional training, has not, however, entirely eliminated the problem.

Several other methodologies for programming generalization across settings have been reviewed by Stokes and Baer (1976). Most, unfortunately, require additional training time, instructional personnel, and/or materials. As such, they entail the further expenditure of time and resources which already exist in short supply in classrooms for severely handicapped students. The obvious practical alternative to this conflict between student needs and available resources is the development of an instructional methodology that would facilitate skill acquisition and generalization across settings within the confines of a single training session, without employing materials or personnel other than those required to achieve initial skill acquisition.

One teaching technique that might accomplish this is the use of distributed practice. Several studies have shown that distributed practice of a skill results in faster acquisition in nonhandicapped students (Adams, 1952; Cook & Hilgard, 1949; Hovland, 1940; Underwood & Goad, 1951) and handicapped students (Baumeister & Berry, 1976; Chasey, 1976; Dunlap & Koegel, 1980), than is achieved under massed practice. This disparity in performance levels of subjects exposed to distributed and massed practice has been explained in terms of the greater motivational effect of varying stimulus presentation (Baumeister & Berry, 1976; Dunlap & Koegel, 1980; Holvoet, Guess, Mulligan & Brown, 1980) and the fatiguing and inhibitory effects of repetitious practice (Adams, 1952; Baumeister & Berry, 1976). If the factor responsible for this effect is motivation or fatigue reduction, other positive effects of distributed practice training might be anticipated. The relationship of nontraining settings has not been tested. It is, however, postulated that a training methodology that increases motivation and decreases fatigue in a training setting may result in superior performances in nontraining environments.

This study assessed the effects of massed practice and distributed practice training on the generalization, across settings, of vocational and daily-living skills taught to adolescent severely handicapped students.

METHOD

Subjects

Two students attending special public school classes for the severely-multiply handicapped served as subjects in this study. These students were chosen because they exhibited no sensory or motor deficits that would inhibit performance on vocational or daily-living tasks.

Student 1 (Roger), a 15-year-old male, was diagnosed as severely retarded and behaviorally handicapped. This student exhibited a moderate degree of stereotyped behavior, such as body rocking and making repetitive noises. Roger was imitative and appeared to have a fairly good receptive vocabulary based on his ability to follow instructions. His expressive vocabulary, however, was limited to echolalic speech.

Student 2 (Janet), a 16-year-old female, was also classified as severely retarded and suffered a minor deficit in auditory acuity. This student exhibited restlessness during instructional sessions, but not to a degree that could be called hyperactive. Her expressive vocabulary was limited to six spontaneous manual signs. Janet was very imitative and, like Roger, appeared to have a fairly good receptive vocabulary. Both students attended the same special class and both had participated as subjects in other experiments earlier in the school year.

Setting

Instruction on the experimental tasks took place in the students' regular classroom, located in a public school for special populations. The arrangement of classroom furniture and equipment was not altered for the purposes of this investigation. The nature of the tasks required the teachers and students to move from place to place within the classroom during the instructional period (e.g., a window-washing task was taught at the classroom windows, and towel-folding was taught at a long table in the middle of the room). Instruction took place during regular class hours so other staff and severely handicapped students were working in the room during the instructional sessions.

Probe (generalization) sessions, which will be discussed in a later section, were conducted in two settings located within the students' school building. Four tasks (disassembling an electric meter, pattern packaging, picture-object matching, and towel folding), were observed in the school vocational center, which was a large room containing nine work tables and several large pieces of machinery. Vocational classes for moderately and severely retarded adolescents were occasionally in session during the time probes were conducted, and lathes and drill presses were sometimes in use. Two tasks (window-washing and dusting) were observed in the school's daily-living area. This was a small room containing bedroom furniture and two easily accessible windows. This room also was occasionally in use by moderately and severely retarded students engaged in bed-making or vacuuming programs.

Tasks and Instructional Objectives

Six tasks were chosen from the prevocational/daily-living skill domain for this study. Three of the tasks chosen were labeled Familiar Tasks since the students had had prior instruction in each of these tasks during the previous semester.

The three Familiar tasks were:

1) Dusting a small table/desk using a dustcloth and spray furniture polish;

2) Packaging dress patterns in groups of four; and

3) Picture-object matching of bolts and washers.

The remaining three tasks were designated Unfamiliar Tasks. Neither student's records indicated any prior experience with nor instruction in these tasks. The Unfamiliar tasks were:

1) Folding bath towels;

2) Removing five screws from an electric meter; and

3) Washing a window using a cloth and a spray window-cleaner.

Instructional objectives, task analyses of 5-7 steps, reinforcement procedures and prompting procedures were devised for each of the six instructional programs. A least-prompts prompting procedure (Lent & McLean, 1976) was used with all six tasks.

The six tasks were randomly paired. Each task was assigned a number from one to six, and a random numbers table was used to pair the tasks (regardless of Familiar/Unfamiliar designation). The first two task numbers selected became the first pair, the second two became the second pair, etc. The pairs of tasks were the same for both subjects. These were:

Pair 1: Disassembling an electric meter and Pattern packaging;

Pair 2: Picture-object matching and Dusting; and

Pair 3: Window-washing and Towel folding.

This resulted in Pair 1 being composed of one Familiar and one Unfamiliar task; Pair 2 being composed of two Familiar tasks; and Pair 3 being composed of two Unfamiliar tasks.

Experimental Procedure

There were two different components to this study: 1) an Instructional component where the students were taught the tasks; and 2) a

Probe component where generalization was tested by giving the students an opportunity to demonstrate the six skills in a different setting with a person other than the teacher. The Instructional and Probe components differed in regard to setting, personnel, time of day, reinforcement for correct responding, and use of a prompting procedure. This study was somewhat complex because the independent variable was the type of practice (massed vs. distributed) that occurred in the Instructional component of the study; whereas, the dependent variable was the degree of generalization (ability to do the task or portions of the task independently and without reinforcement) seen in the Probe component of the study. Both the Instructional and Probe components were conducted on a 1:1 basis.

Massed Practice Instructional Sessions. The Massed practice instructional sessions were conducted daily in the classroom setting by one of the teaching assistants who regularly worked in the students' classroom. The program pairs being taught under this condition were taught from 9:00-10:30 each day. The massed practice instructional session closely resembled those sessions typically conducted in classrooms for severely handicapped students. A trial consisted of the teacher giving the student the needed materials and an initial cue, and then prompting the student when necessary to achieve a correct response. A trial was considered to begin when the initial cue was given and to have ended when the student had correctly completed all the subcomponents (as shown in the task analysis) of the task, either independently or through prompting. The students were given edibles for correct performance of the task even if this performance had been prompted. Ten trials of each of the six programs comprised a session.

Instruction on a task was considered to be in the Massed practice condition when ten repeated trials of the same program occurred so closely together that no other behavior could be expected to be emited between the trials (Mulligan, Guess, Holvoet, & Brown, 1980). This meant that ten trials (each of which consisted of 5-7 subcomponents) of one task (e.g., Pattern packaging) were presented before trials of the next task (e.g., Disassembling an electric meter) were presented.

Programs making up a pair of tasks (e.g., Pair 1: Disassembling an electric meter and Pattern packaging) were always taught in the same consecutive order. Pairs of tasks, however, were presented in all possible orders across sessions. For example, on Monday, the order might be pairs 3, 2, 1; on Tuesday, 1, 2, 3; and on Wednesday, 2, 1, 3. No data are presented in this paper concerning acquisition under this instructional condition.[1]

Distributed Practice Instructional Sessions. The distributed practice instructional sessions were conducted in the classrom setting by the same teaching assistant who had conducted the Massed practice instructional sessions. Two distributed practice sessions, each consisting of five trials of each of the tasks, were conducted daily at 9:00-10:00 and from 1:00-2:00. This two-session format allowed distributed practice across the school day. A trial began with the presentation of materials and ended when the student had correctly completed all the subcomponents of a task either independently or with prompting. Edibles for correct responding were used in these sessions.

Instruction on a pair of tasks was considered to be in the Distributed practice condition if a trial of the second program occurred between every two repeated trials of the first program (Mulligan et al., 1980), and if two daily sessions were conducted with instruction on other tasks occurring between sessions. This means that in the Distributed condition, trials from two programs making up a pair were alternated (e.g., Disassembling an electric meter, Pattern packaging, Disassembling an electric meter, etc.) until five trials of each program had been completed. A second session was then undertaken later in the day.

Programs making up a pair of tasks were always taught in the same consecutive order. The pairs of tasks, however, were conducted in all possible orders across sessions. No data are presented in this paper concerning acquisition under this instructional condition.[1]

Probe Sessions. There was no procedural difference between the probe sessions conducted after Massed Practice Instructional sessions and those conducted after Distributed Practice Instructional sessions. The probe sessions were conducted daily in a nonclassroom setting by one of the investigators who had not previously worked directly with these students in the classroom. The probe sessions were conducted from 11:30-11:45. Each session consisted of one unreinforced trial of each task. The experimenter presented the materials for each task and gave a general cue, such as, "Please fold the towel." The student was given 90 seconds to complete each task. No prompts were given during this time. After 90 seconds, the student was told, "Thank you," regardless of the correctness of his/her performance. No other consequences were provided.

In the Probe sessions the order of the tasks within each pair was the same as it had been during the instructional component. For example, if in Pair 1, Disassembling an electric meter had been taught before Packaging patterns, it was also probed in this order. The three pairs of tasks, however, were presented in all possible orders across sessions. Special care was taken to be sure that the order of the pairs in the Probe session was not the same order used in the Instructional session(s) conducted that same day.

Scoring

In both the Instructional and Probe sessions, the tasks were task analyzed into 5-7 subcomponents. A student was judged to have completed a subcomponent of a task correctly (+) if that subcomponent was completed without verbal, gestural, or physical prompting. If prompting was necessary, a minus (-) was scored. Session scores were computed by counting the number of correct responses and dividing this figure by the number of trials (10 in the Instructional component; 1 in the Probe component) multiplied by the number of subcomponents in the task. This resulted in a figure representing the percent of correct responses made in a session.

Experimental Design

A variation of the multiple baseline design across behaviors was used with both subjects. The dependent variable, generalization, was measured during the Probe sessions. In the multiple baseline design across behaviors, a period of no intervention (baseline) is followed by a period of intervention on a series of tasks or behaviors. The intervention is introduced to each task at a different point in time. In this study, the baseline was that period of time in which a pair of tasks was being taught using massed practice. This was followed by a period in which the experimental intervention (distributed practice) was applied. The introduction of this intervention was lagged across time for each pair of tasks.

Each pair of tasks was assigned a number and the random numbers table was used to determine the order in which the pairs would change from the massed to the distributed condition. This was done seaprately for each subject. The order in which the pairs changed for Roger was: Pair 2, Pair 3, and Pair 1. The order in which the pairs changed for Janet was Pair 3, Pair 1, and Pair 2.

The number of days making up the massed practice baseline was predetermined; however, due to student and staff illness, the distributed condition was shorter than had been planned by five days. This particularly affected the final pair of tasks for each subject, where only three days were obtained in the distributed condition. The number of days each task was taught in each condition is shown in Table 1.

Insert Table 1 about here

Reliability

Interobserver reliability measures were obtained at least once for each student in each condition. Observers were positioned in such a manner as to provide a clear view of the student's performance while preventing observation of the trainer's data sheet. No standard position was specified to the observers as the variety of tasks being taught required them to move about the room frequently.

In both Probe and Instructional settings, occurrence/nonoccurrence reliability was computed using the following formula:

$$\frac{\text{Number of agreements}}{\text{Number of trials x subcomponents}} \text{ X } 100.$$

Table 1

Number of Days Each Instructional Program
Pair was Taught in the Distributed
and Massed Condition

Student	Program Pair	Number of Days in Massed Practice (Baseline)	Number of Days in Distributed Practice
1 (Roger)	2	6 days	16 days
	3	11 days	11 days
	1	18 days	3 days
2 (Janet)	3	8 days	13 days
	1	15 days	6 days
	2	18 days	3 days

RESULTS

Reliability

Interobserver reliability scores are summarized in Table 2. Reliability measures taken during the Instructional component of the study ranged from 76% to 100% for Roger and from 83% to 97% for Janet.

Insert Table 2 about here

Reliability measures taken during the Probe component of the study ranged from 82%-100% for Roger and 80%-100% for Janet.

Generalization Probe Results

The observations of generalization carried out in the Probe component of the study are presented in graphic form in Figures 1 (Roger) and 2 (Janet). The graphs of the Unfamiliar tasks are shown in Column 1 and the graphs of the Familiar tasks are shown in Column 2 of these figures. The two experimental conditions (Massed and Distributed practice) are separated on each graphic display by a vertical dotted line. The solid lines connecting the data points indicate the student's daily performance on a task. Least-squares best fit lines for less than 20 data points (Baer, Note 1) were calculated for each condition to indicate data trends (slopes) and are shown as dotted lines on the graphs. The first through the last data points for each condition were used to determine these trends. The number in parentheses in each condition indicates the slope of the least-squares best fit line, with larger numbers indicating steeper slopes. A negative number indicates a decreasing trend, whereas a positive number indicates an increasing trend.

Table 2

Mean Reliability Scores on Student Performance during Generalization Probes

Program	Condition	$\bar{X}$ Reliability Score for Roger	$\bar{X}$ Reliability Score for Janet
1. Dusting	Massed	87	85
	Distributed	93	97
2. Pattern packaging	Massed	99	90
	Distributed	97	87
3. Picture-object matching	Massed	93	86
	Distributed	98	94
4. Towel folding	Massed	92	96
	Distributed	94	94
5. Disassembling an electric meter	Massed	86	88
	Distributed	96	92
6. Window washing	Massed	80	81
	Distributed	94	85

Student 1 (Roger). Analysis of the slopes of the six graphs in Figure 1 indicates a more rapid rate of generalization was obtained during the distributed practice condition than during the massed practice condition for three of the tasks (Dusting, Window-washing, and Towel folding). Levels close to 100%, a factor that artificially depresses slope because of a ceiling effect, were obtained in the Distributed condition for two other programs (Disassembling an electric meter and Picture-object matching). Only one program (Pattern packaging), a Familiar task, definitively showed more rapid generalization during the massed condition than during the distributed condition. The mean scores also were higher during the Distributed condition for all six programs, as can be seen in Table 3.

Insert Table 3 about here

Visual analyses, however, show that the actual data obtained were very similar in both conditions, with very little change seen at the point where the conditions changed. With the possible exception of the Dusting program and the Disassembling an electric meter program, the data in the Distributed condition appear to be a continuation of the data seen in the Massed condition.

Insert Figures 1 & 2 about here

Student 2 (Janet). Analysis of slopes in Figure 2 shows that Janet showed higher rates of generalization in the Distributed condition than she did during the Massed condition on three of the six tasks (Window-washing, Disassembling an electric meter, and Dusting). The other three tasks favored the Massed condition with two of the Unfamiliar tasks

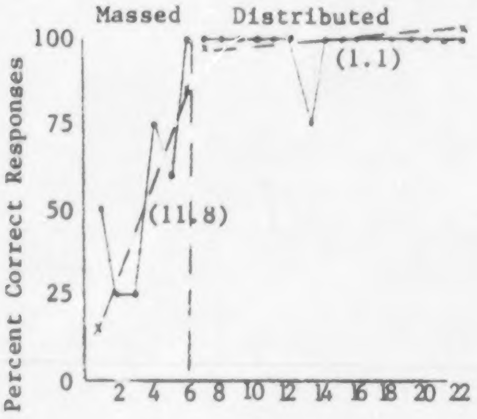
Table 3

Mean Percent Generalization Scores Obtained by Students Under Massed Practice and Distributed Practice Conditions

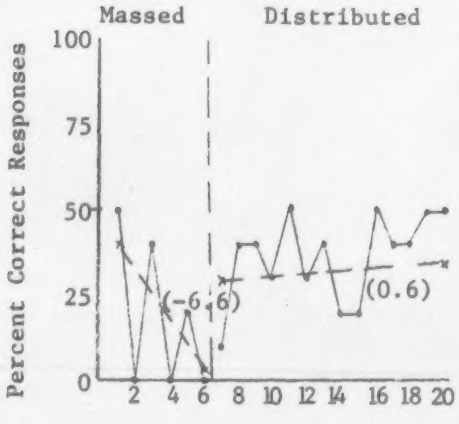
Student	Program	Designation	Generalization Under Massed Cndtn.	Generalization Under Distributed Condtn.
Roger	Dusting	Familiar	18.3%	33.8%
	Pattern Packaging	Familiar	85.7%	95.6%
	Picture-object matching	Familiar	56.3%	98.5%
	Towel folding	Unfamiliar	12.7%	20.5%
	Disassembling an electric meter	Unfamiliar	45.2%	100%
	Window washing	Unfamiliar	30.3%	41.4%
Janet	Dusting	Familiar	31.7%	35.0%
	Pattern Packaging	Familiar	42.5%	69.1%
	Picture-object matching	Familiar	37.2%	62.6%
	Towel folding	Unfamiliar	26.3%	49.4%
	Disassembling an electric meter	Unfamiliar	34.6%	20.0%
	Window washing	Unfamiliar	10.1%	12.5%

Figure 1. The percent of correct responses during the Probe condition made by Roger on six tasks under massed and distributed conditions. Dotted lines indicated least squares best-fit lines and figures in parentheses describe direction and degree of slope.

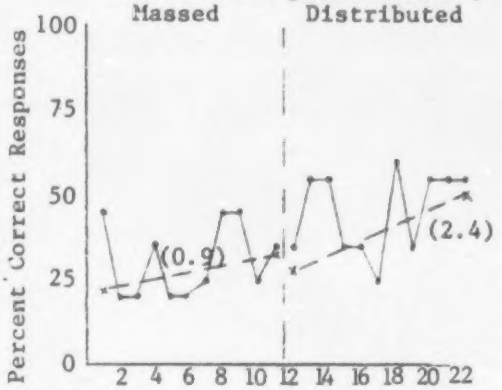
Picture-object matching (familiar)



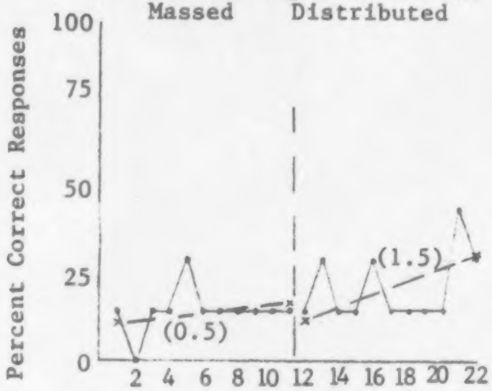
Dusting (Familiar)



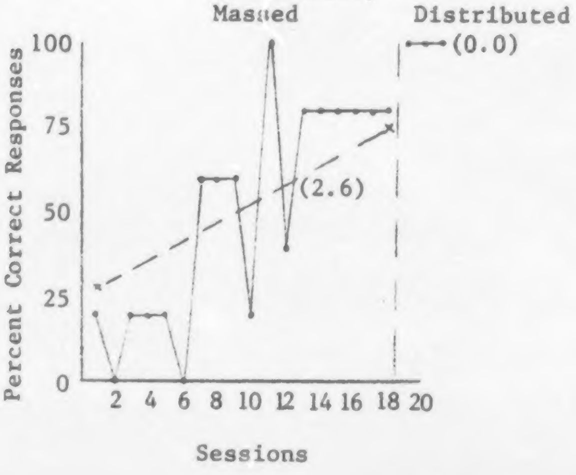
Window-washing (Unfamiliar)



Towel folding (Unfamiliar)



Disassembling an electric meter
(Unfamiliar)



Pattern packaging (Familiar)

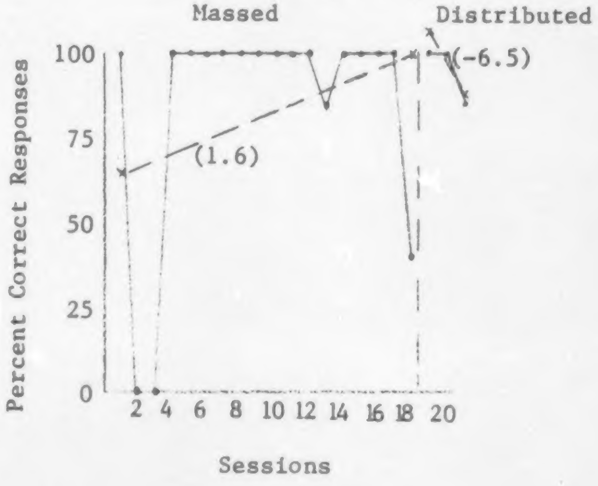
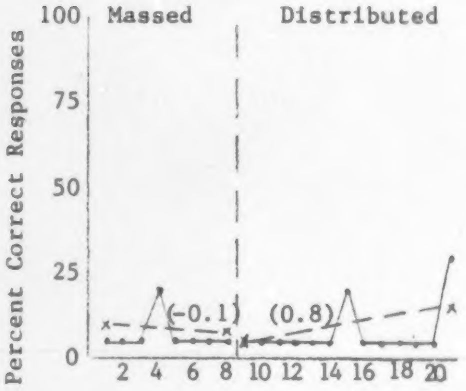
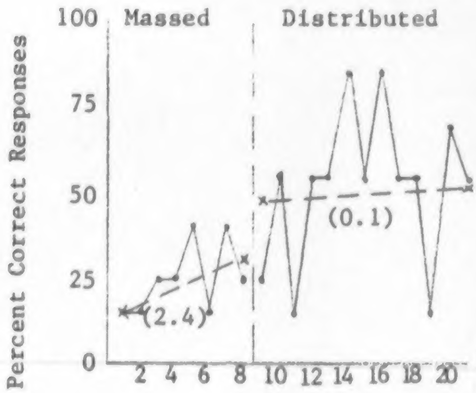


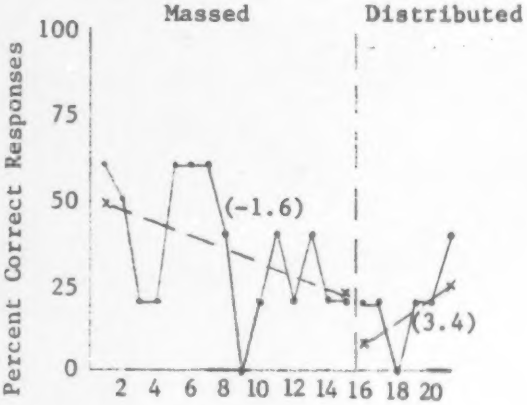
Figure 2. The percent of correct responses during the Probe condition made by Janet on six tasks under massed and distributed conditions. Dotted lines indicate least squares best fit lines and figures in parentheses describe direction and degree of slope.



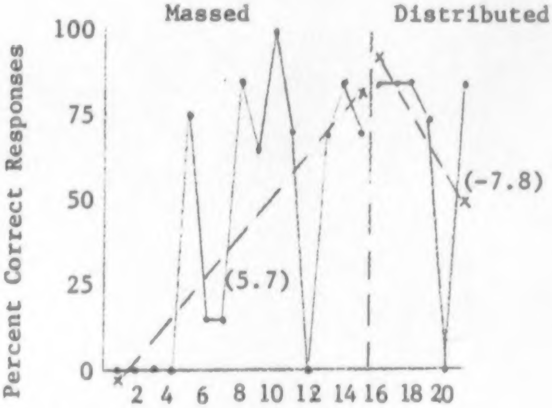
Window washing (Unfamiliar)



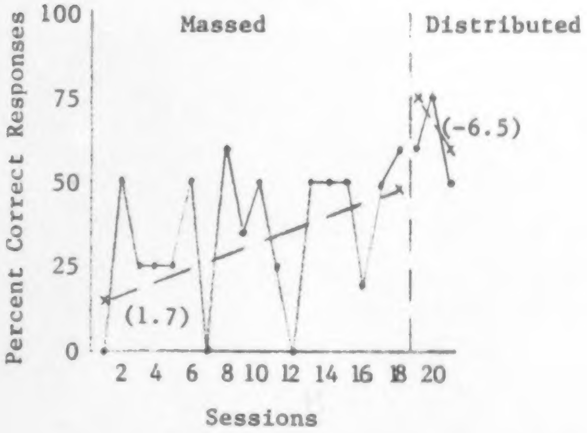
Towel folding (Unfamiliar)



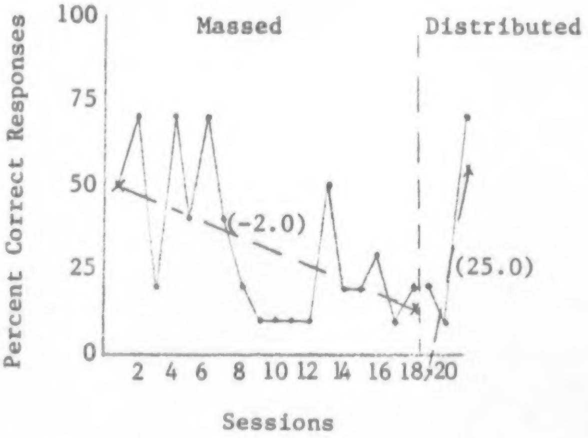
Disassembling an electric meter (Unfamiliar)



Pattern Packaging (Familiar)



Picture-object matching (Familiar)



Dusting (Familiar)

(Pattern packaging and Picture-object matching) actually changing from positive to negative slopes. The means, however, were higher in the Distributed condition for five of the six tasks, as can be seen in Table 3. As with Roger, visual analysis indicated that the actual data obtained in the Distributed condition were very similar to those obtained in the Massed condition and, in fact, could be characterized as simply a continuation of the data established during the Massed condition.

DISCUSSION

This study assessed the effects of Distributed practice instruction on the cross-setting generalization of two severely retarded adolescents. Both students demonstrated higher generalization slopes during distributed practice than they had during massed practice for three of the six tasks. In Roger's case, an argument could be made that superior generalization also occurred on two other tasks since 100% performance was attained and maintained throughout the distributed condition, a factor which could artifically depress the slope. In addition, mean performance increased during the distributed practice condition for all tasks performed by Roger and 5 of the 6 tasks performed by Janet. These data must, however, be interpreted very cautiously since the actual differences between the data are small. Since there was no large change in performance at the times the programs were switched from the Massed to the Distributed practice condition, there is no clear demonstration that the results were not simply the result of maturation or learning. In fact, the similarity of the data, regardless of condition, seems to argue that the statistical differences are largely artificial.

There appeared to be no difference in the effects of Distributed practice on Familiar tasks vs. Unfamiliar tasks.

These data also showed that Massed practice and Distributed practice are not teaching techniques which routinely result in good generalization to other settings. Roger reached high (90% or better) generalization levels on two of the six programs during Massed practice instruction. He was able to maintain this level on these two programs during distributed practice, and to attain a high level on a third program. His other three programs, however, hovered around the 20-40% level which cannot be considered adequate stimulus generalization for independent functioning. Janet's data were even more less encouraging, with no program of the six showing high level generalization in either condition. Her performance seemed to routinely fall between 30% and 60%, which appears to be functionally inadequate.

These data seem to indicate that the teaching practices currently available (Massed practice and Distributed practice) are not sufficient for achieving stimulus generalization in severely handicapped adolescents. Thus, it would appear that the generalization techniques spelled out by Stokes and Baer (1976) and Brown et al. (1976) combined with functional teaching are the best tools currently available for achieving stimulus generalization in spite of their costs in time and resources.

Footnotes

1. These data are available upon request from D. M. Brewer, 2917 Gunnison Trail, Apt. 1068, Fort Worth, TX 76116

Reference Notes

Baer, D. M. An algorithm for calculating the least-squares best fit line for a sequence of 20 or fewer data points. Lawrence, KS: The University of Kansas, Department of Human Development and Family Life, 1977.

REFERENCE

Adams, J. A. Warm-up decrement in performance on the pursuit-rotor. *American Journal of Psychology*, 1952, *65*, 404-414.

Allen, G. J. Case study: Implementation of behavior modification in summer camp settings. *Behavior Therapy*, 1973, *4*, 570-575.

Baumeister, A. A., & Berry, F. M. Distribution of practice and specificity of learning in normals and retardates. *American Journal of Mental Deficiency*, 1976, *72*(2), 227-231.

Brown, L., Branston, M. B., Hamre-Nietupski, S., Pumpian, I., Certo, N. & Gruenewald, L. A. A strategy for developing chronological age appropriate & functional curricular content for severely handicapped adolescents and young adults. *Journal of Special Education*, 1979, *13*, 81-90.

Brown, L., Nietupski, J. & Hamre-Nietupski, S. Criterion of ultimate functioning. In M.A. Thomas (Ed.), *Hey, don't forget about me*. Reston, Va: Council for Exceptional Children, 1976.

Chasey, W. C. Distribution of practice effects on learning retention and relearning by retarded boys. *Perceptual and Motor Skills*, 1976, *43*, 159-164.

Clark, M.G. *Career education for the handicapped child in the elementary classroom.* Denver: Love Publishing Company, 1979.

Cook, B. S. & Hilgard, E. R. Distributed practice in motor learning: Progressively increasing and decreasing rests. *Journal of Experimental Psychology*, 1949, *39*, 172-196.

Dunlap, G., & Koegel, R. L. Motivating autistic children through stimulus variation. *Journal of Applied Behavior Analysis*, 1980, *13*, 619-627.

Griffiths, H., & Craighead, W. E. Gneralization in operant speech therapy for misarticulation. *Journal of Speech and Hearing Disorders*, 1972, *37*, 485-494.

Guess, D., Horner, R., Utley, B., Holvoet, J., Maxon, D., Tucker, D., & Warren, S., A functional curriculum sequencing model for teaching the severely handicapped. *AAESPH Review*, 1978, *3*, 202-215.

Holvoet, J., Guess, D., Mulligan, M., & Brown F. The individualized curriculum sequencing model (II): A teaching strategy for severely handicapped students. Journal of the Association for the Severely Handicapped, 1980, 5, 337-352.

Hovland, I.C. Experimental studies in rote-learning theory, VII. Distribution of practice with varying lengths of lists. Journal of Experimental Psychology, 1940, 27, 271-284.

Lent, J. R., & McLean, B.M. The trainable retarded: The technology of teaching. In N.G. Haring & R.L. Schiefelbusch (Eds.) Teaching special children. NY: McGiaw-Hill, 1976.

Mulligan, M., Guess, D., Holvoet, J. & Brown, F. The individualized curriculum sequencing model (I): Implications from research on massed, distributed, or spaced trial training. Journal of the Association for the Severely Handicapped, 1980, 5, 325-336.

Stokes, F. T., & Baer, M. D. An implicit technology of generalization. Journal of Applied Behavior Analysis, 1976, 10(2), 349-367.

Underwood, B. J. & Goad, D. Studies of distributed practice: I. The influence of intralist similarity in serial learning. Journal of Experimental Psychology, 1951, 42, 125-134.

Walker, H. M., & Buckley, N.K. The use of positive reinforcement in conditioning attending behavior. Journal of Applied Behavior Analysis, 1972, 5, 245-250.

Warren, S. F., Rogers-Warren, A., Baer, D. M., & Guess, D. Assessment and facilitation of language generalization. In W. Sailor, B. Wilcox & L. Brown (Eds.) Methods of instruction for severely handicapped students. Baltimore, MD: Paul H. Brookes, 1980.

Wilcox, B., & Bellamy, G. T. Design of high school programs for severely handicapped students. Baltimore, MD: Paul H. Brookes, 1982.

Evaluation of the ICS Model on Imitation and Generalization

The theories underlying the Individualized Curriculum Sequencing Model suggested that the model might enhance initiation (as evidenced by short response latencies) and generalization because of the use of natural sequences and multiple-exemplar training. The two studies contained in this section examined this premise.

The first study compared response initiation under ICS and massed trial training. Little difference was seen in this variable between the two conditions. The ICS was, however, much more effective than massed trial training in reducing refusals to respond.

The second study looked at the effect of multiple-exemplar training on generalization across items. The results indicated that multiple exemplar training results in complete or partial generalization across items which require similar response topographics.

Initiation Responses of Persons With Severely Handicaps During Individualized Curriculum Sequence (ICS) Training

by

Lynette S. Lacy M.S. Ed.
U.M.K.C. UAF Preschool
Kansas City, Missouri

Marilyn Mulligan Ph.D.

Jennifer F. Holvoet Ph.D.

Doug Guess Ed.D.

University of Kansas
Lawrence, Kansas
1983

This paper was prepared as part of the work pursuant to the "Kansas Individualized Curriculum Sequencing Project" funded as a Handicapped Children's Model Project, U.S. Office of Education (G00-800-1721) from 9-1-80 to 9-1-83.

Abstract

Initiation Responses of Persons With Severely Handicaps During Individualized Curriculum Sequence (ICS) Training

This study compared the responses of two "prompt-dependent" students who were taught using the Individualized Curriculum Sequencing (ICS) Model to the responses obtained when a massed trial teaching procedure was used. Both students reached criterion on initiation responses and achieved very low response latencies by the end of the study on two of the three instructional programs; however, no clear-cut difference in performance between the two training methods was seen on these variables. The ICS was, however, much more effective than massed trial training in reducing refusals to respond.

The question of which curricular strategies should be used to teach severely handicapped persons to become an integral part of society has become of specific concern to special education professionals. Special educators have been given the responsibility of "normalizing" the behavior of severely handicapped persons so they may be assimilated into the mainstream of society. The importance of the handicapped person's independence to this "normalization" process is stressed by Brown, Nietupski, and Hamre-Nietupski (1976). These authors emphasized teaching those behaviors that will enable the severely handicapped person to function as productively and independently as possible in their community. Independent functioning, however, seems not only to be related to skill acquisition, but to spontaneous initiation of the learned skills. In order to achieve a more normalized existence, severely handicapped persons must initiate appropriate responses in their environment, just as nonhandicapped persons demonstrate their independence in their daily environment(s) by initiating responses. There are very few data specifically discussing whether initiation is a skill that is generally lacking in the severely handicapped population. There are, however, studies in related areas that seem to substantiate such a claim.

Several laboratory studies have compared response latencies (i.e., the amount of time that it takes from the presentation of a stimulus until the initiation of a response) in handicapped and nonhandicapped individuals (Baumeister, Hawkins, & Kellas, 1965; Berkson, 1960a; 1960b; Berkson & Baumeister, 1967; Wade, Newell, & Wallace, 1978). These studies have consistently found that handicapped persons have longer response latencies than nonhandicapped persons. Berkson's studies

(1960a; 1960b) also indicated that there was a strong correlation between level of intelligence and response latency. These data showed that the lower the measured intelligence, the longer the response latency. Berkson noted that this increase in response latency was not correlated with how fast the subjects visually perceived the stimuli, how fast they were able to make a choice nor how fast they could plan the motor movement they needed to make. Response latency was, however, specifically correlated with a slowness in initiating the planned response. This finding was supported in later studies by Baumeister, et al. (1965) and Wade, et al. (1978).

These results should be viewed with caution, especially in light of the study by Dingman and Silverstein (1964) who reported that when they controlled for motor disabilities there was no correlation between intelligence and response latency. Nonetheless, these data and the general finding that generalization across settings, persons, and materials is often deficient and needs to be taught (Koegel, Egel, & Dunlap, 1980; Warren, Rogers-Warren, Baer, & Guess, 1980) to severely handicapped individuals makes it appear that educators need to be more sensitive to the variables of response latency and speed of response initiation if they wish the student to initiate responses in non-training environments. Furthermore, it is important to analyze current teaching practices to determine which techniques effectively reduce latency, enhance and facilitate initiation.

This study compared the initiation responses of two "prompt-dependent" severely handicapped students taught using the Individualized Curriculum Sequencing Model (Holvoet, Guess, Mulligan, & Brown, 1980) to these repsonses obtained when they were taught using a more traditional massed trial teaching procedure. The Individualized Curriculum Sequencing Model (ICS) is a teaching strategy designed to increase the independence of severely handicapped students in their home and school environments. Aspects of the ICS that might facilitate self-initiation and short response latencies are: 1) functional and age-appropriate materials and tasks (cf. Brown, et al., 1976) 2) training in functional settings (cf. Guess, Horner, Utley, Holvoet, Maxon, Tucker, & Warren, 1978; Sailor & Guess, 1983); 3) use of skill clusters (cf. Holvoet, et al., 1981); 4) concurrent task sequencing (cf. Panyan & Hall, 1978; Schroeder & Baer, 1972); and 5) distributed practice (cf. Mulligan, Guess, Holvoet, & Brown, 1980; Mulligan, Lacy, & Guess, 1982).

Massed trial training, a more traditional classroom teaching procedure, teaches only one skill during an instructional period and has the student practice the skill several (e.g., 10-40) times in rapid succession during that time. Thus, massed trial training might enhance initiation by allowing the student repeated practice and feedback on the same task in a short period of time. Massed trial training also can include appropriate and functional materials and tasks.

METHOD

Subjects

Two severely handicapped male adolescents participated in the study. The subjects were enrolled in a public school self-contained classroom with seven other severely handicapped students. The students were chosen to participate in the study because their teacher reported these students rarely spontaneously initiated responses during structured or unstructured classroom time. Both these students showed good direction following skills, but always seemed to wait for an instruction from an adult instead of trying to communicate, move, or obtain items on their own.

The first subject, Tom, was sixteen years old and lived at home with his parents. He was confined to a wheelchair which he was able to push by himself upon request. He was classified as nonverbal, but responded to simple verbal directions and was able to use thirty line-drawn picture symbols on a communication board upon request. Tom was assessed using the TARC Assessment Inventory for Severely Handicapped Children (Sailor & Mix, 1975). His total score was in the 25th percentile.

The second subject, Bill, was seventeen years old and resided in a state institution. He was nonverbal but responded to two- and three-part directions that were spoken and manually signed. Bill had an expressive sign vocabulary of approximately 10-15 nouns, and one verb. He would use this sign language when a known object was placed in front of him and an adult asked/signed, "What want?" This student was also assessed using the TARC Assessment Inventory for Severely Handicapped Children. His total score was in the 60th percentile.

Setting

The study was conducted in a wide variety of settings within the students' public high school building. This school served only special populations. The first setting was the students' classroom. This was a typical classroom for severely handicapped students. It included two large tables, two or three individual student desks and chairs, a teacher's desk, several shelves, a large floor mat, and a beanbag chair. The second setting was the student cafeteria which consisted of a cafteria serving area, a dirty dish return, and several large tables where staff and students from several classrooms ate. The third area was a large simulated workshop area with several workbenches, tables, and shop equipment such as drill presses and saws. The students' teacher and seven classmates were usually present in the workshop area while the study was being conducted there. A fourth area was the boy's restroom which consisted of one sink, a mirror, two urinals, one nonadapted toilet, one adapted toilet, and several shelves. Usually the experimenter and student were the only persons in the restroom during the time the study was being conducted there. The fifth training area was a home economics area consisting of a kitchen and a bathroom. The kitchen was furnished with two stoves, four sink/counter/cabinet areas, a kitchen table, and a washing machine and dryer. The bedroom was a small carpeted room with a bedside table, a small lamp, a twin bed, and a wooden chest of drawers. The students' teacher and classmates were present in the home economics area during the study. A sixth area used for training was a regulation size gymnasium with a marked basketball court, bleachers, and climbing equipment. Other gross motor equipment was sometimes present in the gymnasium. The students' classmates and teacher were usually present in the gymnasium while the study was being conducted there. The last training area, the hallways of the school, were dimly lit, very long (approximately ½ city block), and lined with student lockers and the doors to the various classrooms. The hallways were usually occupied by several staff and/or students moving between classes or from the bus area.

Instructional Programs/Materials

Three target behaviors for each student were chosen for intervention by the experimenter and the classroom teacher, based on the student's assessed needs. Instructional programs were then written for each of these target behaviors.

The instructional programs designed for Tom were: 1) initiating a greeting by waving; 2) initiating communication using a communication board, and 3) transporting items.

The program designed to teach a greeting response defined the target behavior as: "Tom will wave 'Hi' by raising one empty hand above his elbow and making at least two back and forth motions, either from the shoulder, the elbow, or the wrist. He should not make contact with any part of his body, another person, or an object." The cue for initiation was the approach of a familiar (seen at least 5 times) person saying "Hi". No materials were needed for this program.

The program designed to teach communication used a communication board made from a laminated manila folder with thirty line drawings on it and several items (soap, cup, toothbrush, food) that were needed or desired by Tom. The communication board was attached to Tom's wheelchair. The other items were placed within the student's view, but out of his reach. The target behavior was defined as: "Tom will tap another person on the arm to gain attention, pick up or pull over his communication board, open it and point to the picture of the desired object." The cue for initiation was bringing the presence of the unobtainable item to Tom's attention through pointing to it or holding the desired item out of his reach.

The program for teaching Tom to transport items used a plastic dishpan with a variety of small objects, such as a toothbrush and hairbrush, in it. The target behavior was defined as: "Tom will pick up the dishpan from the table, put it in his lap and hold it with at least one hand for 10 seconds while being pushed in his wheelchair by another person." The cue was the presence of the dishpan and a verbal cue, "Get your stuff and let's go."

The three instructional programs devised for Bill were: 1) initiating communication using manual sign language; 2) using a self-monitoring tally sheet; and 3) obtaining and returning data sheets from a 3-ring binder.

The communication program used manual signs derived from Signing Exact English (Gustason, Pfetzing, & Zawolkow, 1980) for the following items: cup, food, toothbrush, soap, and a variety of materials used in a prevocational program. These items were placed where they were clearly visible, but could not be freely obtained by the student. The target behavior was defined as, "Bill will correctly sign 'want (item)' within ten seconds of the cue." The cue for initiation was bringing the presence of the unobtainable item to Bill's attention.

The program devised to teach the student to use a self-monitoring tally sheet utilized a plain 3" X 5' piece of blank paper and a pencil. The cue for initiation of the response was correctly completing one response in a prevocational program. Using the self-monitoring tally sheet was defined as, "After making three marks on a tally sheet (one for each correctly completed item), Bill will raise his hand above his shoulder in order to signal an adult to give him an edible reinforcer." The cue for initiation was having three marks on the tally sheet.

Lastly, the program designed to teach him to obtain and return data sheets from a three-ring binder used a plastic 3-ring notebook divided into three sections (lunch, daily living, and workshop) by heavy cardboard dividers that were different colors. Each section contained a data sheet made up of at least two stapled pages. Holes had been punched in each data sheet so they would fit in the notebook. The target response was defined as, "Bill will open the notebook, turn to the correct section, and put in or take out his data sheet." The cue for initiating putting the data sheets in the notebook was being given the notebook and the data sheet and the verbal cue, "We're finished in the __ area." The cue for getting a data sheet out was being given the notebook and the verbal cue, "We are ready to work in the __ area."

The experimenter conducted the teaching trials for all of the programs with the exception of Tom's initiation of a greeting program. In this program, two teaching assistants from Tom's classroom assisted during the experimental condition where ICS training was in effect.

Experimental Procedure

This study evaluated, across three tasks: 1) independent initiation of responses, 2) latency of responses, and 3) refusals to respond after prompting. These variables were compared under traditional massed trial teaching procedures and under ICS teaching procedures. These two teaching procedures differed across three parameters: 1) setting; 2) distribution of teaching trials; and 3) use of skill clusters/concurrent training.

Traditional Massed Trial Teaching Model. The only setting used for instruction in the traditional massed trial teaching condition was the student's classroom. A daily 45 minute instructional period consisted of 10 consecutive trials of the first task, followed by 10 consecutive trials of the second task, followed by 10 consecutive trials of the third task (a massed practice paradigm for each task). The tasks were taught in isolation from other tasks, so there was no concurrent training or use of skill clusters in this condition.

ICS Teaching Model. Instruction in the ICS condition was conducted in a wide variety of settings (cafeteria, workshop, home economics area, gymnasium, and hallway) at the times the behaviors should naturally be expected to occur. For example, a trial of initiation of a greeting was conducted in the hallway when the student arrived at school. It was also taught at other times throughout the school day when the student had an opportunity to enter a different setting where other staff were present. Thus, instructional trials were distributed throughout the school day rather than being massed at one time (cf. Mulligan et al., 1980). Because of this distribution, the instructional tasks that were analyzed in this study were scheduled in skill clusters with other behaviors. For example, after Tom had greeted someone in the workshop area (one of the analyzed skills), he did a communication trial (another analyzed skill), then he did a 2-part prevocational skill (a task that was not analyzed in this study), and completed the cluster by transporting the prevocational materials to a cabinet (the third analyzed skill). This use of skill clusters maximized concurrent training. Ten distributed trials of each task were conducted on a daily basis.

Trials. In this study, a trial was defined as the time that elapsed from the presentation of the cue to the completion of a correct response whether that response was independently achieved or prompted by the experimenter. The trials were structured to provide the student with an opportunity to initiate a response. In both the traditional and the ICS conditions, a trial consisted of the experimenter giving a general verbal/ signed cue or showing an object to a student and waiting for the response. If the student initiated a response within ten seconds, reinforcement was given. If the student responded incorrectly or did not respond at all, the teacher implemented a prompting procedure as specified by the instructional program. The prompting procedure, a least-prompts strategy, was designed to provide the student with as little assistance as was necessary to correctly complete the task.

Measurement

Three measures were taken during both the massed and ICS conditions. These were: 1) whether or not the student initiated a response, 2) the latency of the response, and 3) when a physical prompting procedure was necessary, whether the student resisted the prompting procedure or refused to engage in the response (e.g., tantrumed, struggled, turned away, tried to leave).

Initiation responses were scored as occurring (+) if the student made a response within 10 seconds, whether or not the response was correct. If the student did not initiate a response within 10 seconds of the cue presentation, a minus (-) was recorded and the experimenter began the prompting procedure. The prompting procedure was also implemented if the student initiated an incorrect response or approximated a correct response within the ten second limit. Initiation responses were praised even if they did not result in correct performance. Initiation scores for each session were converted to percent by dividing the number of trials on which the student initiated a response by the total number of trials and multiplying the result by 100.

Latency was recorded as the number of seconds that elapsed between the initial cue from the experimenter (in the form of a verbal direction or object presentation) to the student beginning a response. If the student did not respond within ten seconds, a latency of 11 seconds was recorded and the prompting procedure was implemented. Mean latency scores were calculated for each session by adding the latency scores for each trial and dividing that total by the number of trials.

Refusals to respond to prompting were defined as the student resisting physical prompting through body tightness, tantruming, pushing away materials, trying to turn away from the task, refusing to look at an item, or attempting to leave the session. Refusals to respond to prompting during each session were converted to percent by dividing the number of refusals by the total number of trials, and multiplying the quotient by 100.

Experimental Design

A multiple baseline across tasks design (Cooper, 1974) was used with each subject. This design allows a visual comparison of the responses under ICS training conditions to those obtained under traditional massed trial training procedures. There were two conditions: massed trial training (this served as a baseline measure) and ICS training. Each program was initially implemented in the massed condition. The programs were changed to the ICS condition at varying times when there were sufficient data points to establish a stable trend in massed trial initiation. Table 1 illustrates the number of days each program was in the two conditions. Note that the use of the multiple baseline across

----

Insert Table 1 about here

Table 1

Number of Days Each Instructional Program Was Taught in Massed and ICS Conditions

Student	Program	Number of Days in Massed	Number of Days in ICS
1.(Tom)	Initiating Communication	3	22
	Transporting items	7	18
	Initiating a greeting	18	7
2.(Bill)	Initiating Communication	5	23
	Using a tally sheet	8	15
	Obtaining/Returning data sheets	19	5

behaviors design means that sometimes the student would be experiencing both massed and ICS training on different tasks during a single day. For example, Tom in session 8 was experiencing Initiation of Communication trials and Transporting Items trials in the ICS mode and Initiation of a Greeting in the massed trial condition.

Reliability

Reliability data were taken at least once for all programs in both the traditional massed trial and the ICS conditions. In the massed condition, an observer was seated near the experimenter where he/she was able to see the student performing the response but not within view of the experimenter's data sheet. During the ICS condition, an observer closely followed the teacher and student to the different settings in which the tasks were performed. In both conditions the observer scored the student's behavior in terms of whether the student initiated a response and whether the student refused to respond if a physical prompt was needed. The observer and the experimenter used the same stopwatch to measure the latency of the response, so no reliability data were taken on this measure. The same type of data sheets were used by both the experimenter and observer. The reliability scores were calculated separately for each task. The formula used to determine degree of reliability was:

$$\frac{\text{Number of agreements}}{\text{Number of agreements + disagreements}} \times 100.$$

RESULTS

Reliability

Table 2 shows the mean interobserver reliability scores for both students on each task. The reliability scores for initiation for Tom on Tasks 1, 2, and 3 ranged from 90% to 100%. Bill had mean reliability scores for initiation of 100%.

Insert Table 2 about here

The mean reliability for refusal to respond for Tom was 85% with a range of 0% to 100%. Bill never refused to respond during any reliability session.

Performance of Students

The results for each student are presented in graphic form. Figure 1 illustrates Tom's daily percent scores for the three variables (initiation, latency, and refusal to respond to prompting). These scores are shown separately for each program. The solid lines indicate the student's daily performance calculated as a percent (initiation and refusal to respond to prompting), or mean duration (latency). Least-squares best fit lines (Baer, Note 1) were calculated for each condition to illustrate learning trends and are shown as broken lines on the graphs. The first and last data points for each condition were used to determine these lines. The number in parentheses in each condition indicates the slope of the least-squares best fit line.

Figure 2 illustrates Bill's scores on each of his programs, and is organized in the same way as Figure 1.

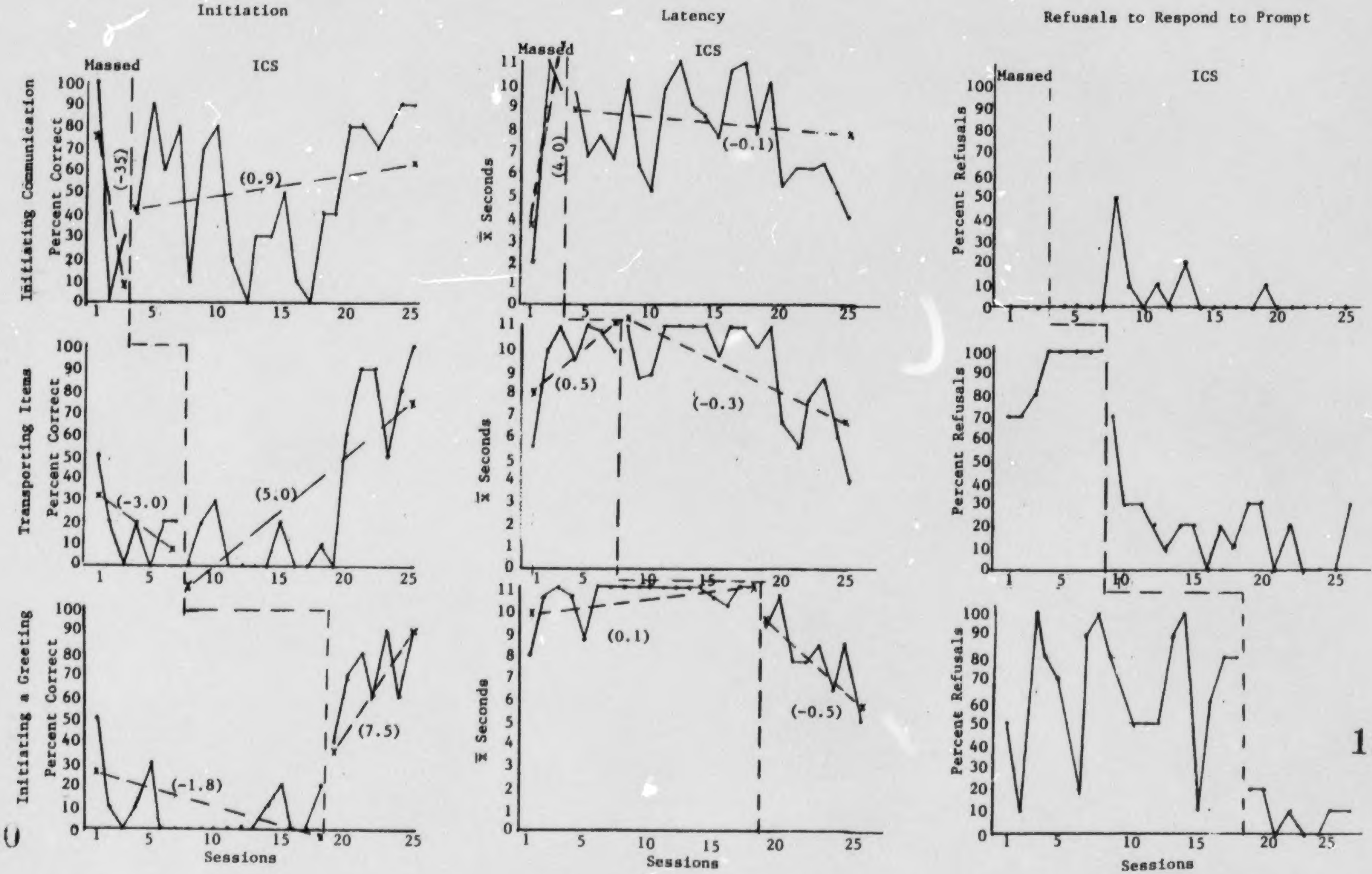
Table 2

Reliability for Initiation Responses
and Refusals to Respond to a Prompt

	Condition	Mean Reliability Scores for Initiation Responses	Mean Reliability Scores for Refusals to Respond to a Prompt
Student 1 (Tom)			
Task 1 (Initiating communication)	Massed	90%	100%
	Functional	100%	95%
Task 2 (Transporting items)	Massed	90%	100%
	Functional	100%	51%
Task 3 (Initiating greeting)	Massed	100%	80%
	Functional	100%	100%
Student 2 (Bill)			
Task 1 (Initiating communication)	Massed	100%	not applicable
	Functional	100%	" "
Task 2 (Obtaining & Returning data)	Massed	100%	" "
	Functional	100%	" "
Task 3 (Using a self-monitoring Tally sheet)	Massed	100%	" "
	Functional	100%	" "

Figure 1

Student 1 (Tom)



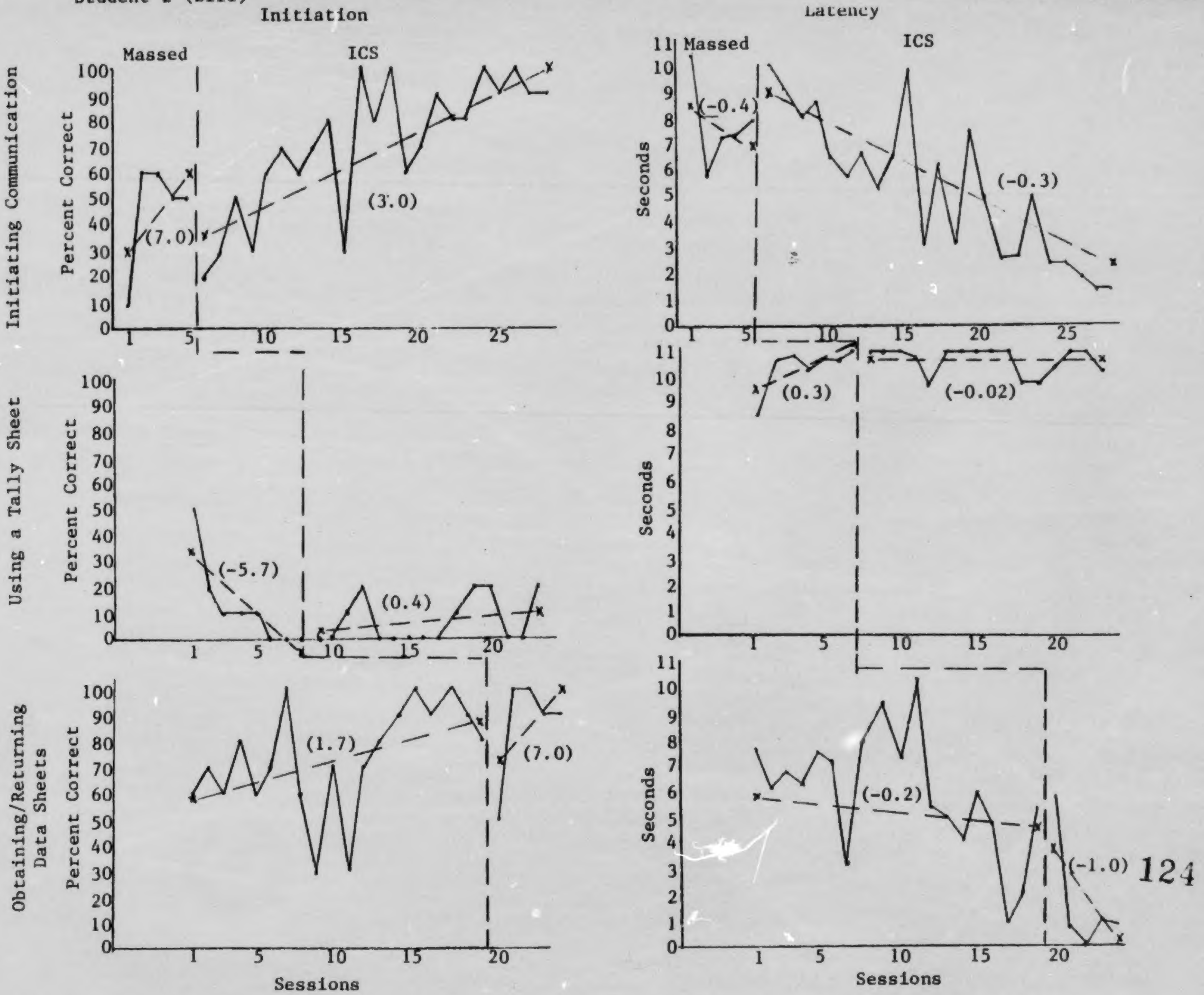
Initiation

Latency

Refusals to Respond to Prompt

Figure 2

Student 2 (Bill)



Student 1 (Tom)

Initiation. Visual analysis of Tom's initiation graphs in column 1 of Figure 1 indicates that there was, on Program 1 (Initiating communication): a) an increase in initiation over massed trial responding for the first 4 sessions of the ICS condition; b) a variable but generally downward trend for the next 10 sessions in the ICS condition; and c) a return to a high level of initiation in session 20 and a generally increasing trend for the rest of the sessions. Criterion (80% or better for 3 out of 5 sessions) was reached on this program for the initiation response.

On Program 2 (Transporting items) it can be seen that: a) no change in initiation occurred during the first 12 sessions of the ICS condition, and b) there was a sudden rise in initiation on session 19 and a generally increasing trend for the rest of the sessions. Initiation performance was at criterion levels by the end of the study.

On Program 3 (Initiating a greeting), it is apparent that: a) there was a substantial increase in initiation over that seen in the massed condition when the ICS condition was begun, and b) this improvement continued for the rest of the study. Initiation responses in this program were close to criterion levels. It is also apparent that the large increases in initiation in the first two programs began at the time the third program was switched to the ICS condition.

Latency of response. The second column of graphs in Figure 1 illustrated Tom's latency of response for all three programs. It should be pointed out that low response latencies are desired in all these programs, and that data points at the top of the graphs (11 seconds) indicate that the student did not initiate the response in the allotted 10 seconds.

In Program 1, it can be seen that: a) a change from the massed condition to the ICS condition resulted in an initial decrease in response latency (5 of the first 7 sessions); b) this initial decrease was followed by a variable, but generally upward trend for the next 9 sessions; and c) in session 20, latency dropped considerably and continued at these low levels.

In Program 2, it appears that: a) there was virtually no change from the massed trials condition in the response latency on the first session of the ICS condition, b) there was a slight declining trend during the second and third ICS session, then a return to previous high levels until session 20; and c) on session 20 (in the ICS condition) there was a substantial drop in response latency which continued in a variable manner until the end of the study.

In Program 3, it can be seen that: a) responses in the first two sessions of the ICS condition were fairly similar to those obtained early in the massed trial condition; and b) beginning in the third ICS session there was a large drop in response latency which began a clear decreasing trend. As was seen in the Initiation data, decreases in response latency were seen in all three programs close to the point where the third program changed from the massed trial to the ICS condition.

Refusal to respond to prompting. Column 3 of Figure 1 illustrates the percent of refusal to respond to a physical prompt on the three programs. It should be remembered that a physical prompt occurred in response to trials where the student did not initiate a response and to those trials where the student made an incorrect initiation. A low score on this variable would be desirable.

On Programs 2 and 3, it can clearly be seen that there was a substantial and continued decrease in refusals to respond at the point the ICS condition was introduced. In Program 1, the refusal rate was at the zero level during the massed condition, and varied from 0%-40% in the ICS condition.

Student 2 (Bill)

Initiation. Column 1 of Figure 2 illustrates Bill's initiation responses in the three programs. It would be desirable to have high scores and/or steep positive slopes on this variable. In Program 1 (Initiating communication), it is apparent that: a) the first three sessions of the ICS condition are characterized by lower initiation scores than had been obtained under massed trial training; and b) there was generally an upward trend in the ICS condition, and criterion was met. The slope in the ICS condition, however, was less steep than that seen in the massed trial condition.

In Program 2 (Using a self-monitoring tally sheet), it can be seen that generally the same level of initiation was obtained in the ICS condition as was obtained in the massed condition. The slope was decreasing under the massed condition and virtually flat under the ICS condition.

In Program 3 (Obtaining and returning data sheets from a 3-ring binder), initiation was less on the first ICS session than it had been in the massed trial condition, but by the second session had returned to the criterion levels that had initially been achieved, and continued at that level. The slope in the ICS condition was slightly steeper than that attained under the massed trial condition.

Latency of response. The second column of graphs in Figure 2 shows Bill's latency of response data for all three programs. If the student is doing well on this variable, the scores should be low and the slopes should be negative.

In Program 1, the latency of response was higher during the first 4 sessions of the ICS condition than it was under the massed trial condition. After this point, however, there is an apparent continuation of the decelerating trend first seen in the massed trial condition and very low response latencies are reached. The slope was negative in both conditions, but was steeper in the ICS condition.

In Program 2, the response latency was at very high levels in both the massed and ICS conditions. The slope was ascending in the massed trial condition and was stable at the 10-11 second level in the ICS condition.

In Program 3, there was an initial rise (1 session) in response latency when the ICS condition was begun, but after this point the downward trend seen in the massed condition was reestablished. Again the slope in the ICS condition was steeper than that seen in the massed condition.

Refusals to respond to prompting. Bill never refused to respond on any of his programs under either condition.

DISCUSSION

This study evaluated the effects of the ICS teaching technique on the initiation, latency of responding, and refusals to respond to prompting of two severely handicapped learners. The results varied for the two students. For Student 1 (Tom), it generally appeared that there were more initiation responses, latency of responding was lower, and there were fewer refusals to respond in the ICS condition than occurred in the massed trials condition. Nonetheless, these data must be interpreted with a great deal of caution. The only variable that showed an immediate change when ICS training was introduced was Refusal to Respond to Prompting, and the change was only seen in two of the three programs. The improvement seen in slope and mean responding on Initiation and Response Latency was almost entirely due to the improvement in all the programs that occurred when the third program was moved into the ICS condition. Prior to that point, the student generally showed a pattern that was more variable, but generally similar to that seen in the massed trial condition. One explanation for this might be that the ICS teaching procedure affected initiation and latency of responding cumulatively. In other words, there might be a minimal number of programs which must be taught using the ICS strategy before the effects are seen. Another possible explanation for this effect is that some other controlling factor in the environment, not identified by the investigators, was responsible for the changes.

The second student (Bill), showed steeper initiation slopes under the ICS conditions on two of the three programs, with the slope for one of these programs (Using self-monitoring tally sheet) changing from negative to positive. Bill also showed steeper latency of response slopes under the ICS condition on two of the three programs. These differences, however, are very small for both the variables and could be interpreted as simply showing a continuation of the learning trend seen under the massed trial condition. In general, the programs that were stable, erratic, or improving during the massed condition remained that way when moved into the ICS condition. No cumulative effect across programs was seen in the second student's Initiation responses nor in his Response Latency. This student never refused to respond to prompting, so this variable showed no difference between conditions.

In spite of the fact that the results did not show a clear-cut effect of the ICS in the training of initiation responses and in the lowering of response latency, it is encouraging that each of the two students reached criterion on initiation and had very low response latencies by the end of the study on two of the three programs. This study substantiated the findings of several authors (Mayhew, Enyart, & Anderson, 1978; Stokes, Baer & Jackson, 1974) that initiation responses could be trained in naturalistic settings.

The most evocative result in the study was the apparent effectiveness of the ICS training strategy in reducing refusals to respond. This result is in accord with the findings of Mulligan, Lacy and Guess (1982) who compared massed and distributed practice, using a group design, and found that distributed practice did not improve acquisition, but did significantly reduce refusals to respond.

Summary

In the present study it was shown that severely handicapped students could be taught to increase initiation responses and to reduce response latencies when an ICS training strategy was used. It was, however, apparent that students are differentially affected by the procedure and that some students (such as Student 1) may need to experience several programs in the ICS mode for it to be effective. It should be pointed out that since these students also were trained under a traditional massed trials system (during the baseline component), it is possible that these effects are due to the combination of ICS and massed trial training. It is also not known whether the same effects could have been achieved if massed trial training had been used alone.

This study also verified that the ICS technique is an effective method of decreasing refusal to respond to prompting and, thus, may be a valuable tool for those students who exhibit this problem.

It seems that much more research needs to be done on the topic of response initiation and response latency. It also appears that techniques are needed that will more consistently and reliably increase initiation across a wide variety of programs.

List of Figures

Figure Number	Caption
1	Tom's daily mean scores for initiation, latency to respond and refusal to respond to prompting on each of three tasks.
2	Bill's daily mean scores for intiation and latency to respond on each of three tasks.

Reference Notes

Baer, D. M. An algorithm for calculating the least-squares best fit line for a sequence of 20 or fewer data points. Lawrence, KS: The University of Kansas, Department of Human Development and Family Life, 1977.

REFERENCES

Baumeister, A. A., Hawkins, W. F., & Kellas, G. Reaction speed as a function of stimulus intensity in normals and retardates. Perceptual Motor Skills, 1965, 20, 649-652.

Berkson, G. An analysis of reaction time in normal and mentally deficient young men. II. Variation of complexity in reaction time tasks. Journal of Mental Deficiency Research, 1960a, 4, 69-77.

Berkson, G. An analysis of reaction time in normal and mentally deficient young men. III. Variation of stimulus and of response complexity. Journal of Mental Deficiency Research, 1960b, 4, 69-77.

Berkson, G., & Baumeister, A. A. Reaction time variability of mental defectives and normals. American Journal of Mental Deficiency, 1967, 72(2), 262-266.

Brown, L., Nietupski, J., & Hamre-Nietupski, S. The criterion of ultimate functioning and public school services for severely handicapped students. In M. Thomas (Ed.), Hey, don't forget about me: New directions for serving the severely handicapped. Reston, VA: Council for Exceptional Children, 1976.

Cooper, J. O. Measurement and analysis of behavioral techniques. Columbus, OH: Charles E. Merrill, 1974.

Dingman, H. F., * Silverstein, A. B. Intelligence, motor disabilities, and reaction time in the mentally retarded. Perceptual Motor Skills, 1964, 19, 791-794.

Guess, D., Horner, R. D., Utley, B., Holvoet, J., Maxon, D., Tucker, D., & Warren, S. A functional curriculum sequencing model for teaching the severely handicapped. AAESPH Review, 1978, 3, 202-215.

Gustason, G., Pfetzing, D., & Zawolkow, E. Signing Exact English. Roosmoor, CA: Modern Signs Press, 1972.

Holvoet, J., Guess, D., Mulligan, M., & Brown, F. The individualized curriculum sequencing model (II): A teaching strategy for severely handicapped students. Journal of the Association for Severely Handicapped, 1980, 5, 337-351.

Koegel, R. L., Egel, A. L., & Dunlap, G. Learning characteristics of autistic children. In W. Sailor, B. Wilcox, & L. Brown (Eds.), Methods of instruction for severely handicapped students. Baltimore: Paul H. Brookes, 1980.

Mayhew, G. L., Enyart, P., & Anderson, J. Social reinforcement and the naturally occurring social responses of severely and profoundly retarded adolescents. _American Journal of Mental Deficiency_, 1978, _83_, 164-170.

Mulligan, M., Guess, D., Holvoet, J., & Brown, F. The individualized curriculum sequencing model (I): Implications from research on massed, distributed or spaced trial training. _Journal of the Association for the Severely Handicapped_, 1980, _5_, 325-336.

Mulligan, M., Lacy, L., & Guess, D. The effects of massed, distributed and spaced trial sequencing on severely handicapped students' performance. _Journal of the Association for the Severely Handicapped_, 1982, _7_, 48-61.

Panyan, M. C., & Hall, R. V. Effects of serial versus concurrent task sequencing on acquisition, maintenance and generalization. _Journal of Applied Behavior Analysis_, 1978, _11_, 67-74.

Sailor, W. & Guess, D. _Severely handicapped students: An instructional design_. Boston: Houghton Mifflin, 1983.

Sailor, W. & Mix, B. J. _The TARC assessment system_. Lawrence, KS: H and H Enterprises, 1975.

Schroeder, G., & Baer, D. Effects of concurrent and serial training. _Developmental Psychology_, 1972, _6_, 293-301.

Stokes, T. F., Baer, D. M., & Jackson, R. L. Programming the generalization of a greeting response in four retarded children. _Journal of Applied Behavior Analysis_, 1974, _7_, 599-610.

Wade, M., Newell, K., & Wallace, S. Decision time and movement times as a function of response complexity in retarded persons. _American Journal of Mental Deficiency_, 1978, _83_, 135-144.

Warren, S. F., Rogers-Warren, A., Baer, D. M., & Guess, D. Assessment and facilitation of language generalization. In W. Sailor, B. Wilcox, & L. Brown (Eds.), _Methods of instruction for severely handicapped students_. Baltimore: Paul H. Brookes, 1980.

Effects of Multiple Exemplar Training on Generalization of Prevocational Skills

Blane Brown, B.A.

Jennifer F. Holvoet, Ph.D.

Doug Guess, Ed.D.

University of Kansas
Department of Special Education

1983

This paper was prepared as part of the work pursuant to the "Kansas Individualized Curriculum Sequencing Project" funded as a Handicapped Children's Model Project, U.S. Office of Education (G00-800-1721) from 9-1-80 to 9-1-83.

The ability to apply what has previously been learned to new situations has been termed generalization. Generalization appears to be a natural process in most nonhandicapped persons, but several investigators have documented that students with severely handicapping conditions do not exhibit the skills they have learned across settings, persons, or objects without generalization training (Stokes & Baer, 1977; Wehman, Abramson, & Norman, 1977; Brown & York, 1974). Stokes and Baer (1977) and Drabman, Hammer, and Rosebaum (1979) have outlined several strategies which should produce generalized performance in this population. Unfortunately, most have yet to be systematically tested in terms of efficacy and/or efficiency in classrooms for severely handicapped individuals. One method proposed by Stokes and Baer for improving generalization is training sufficient exemplars. These authors note that "very little research concerned with generalization programming has dealt with the training of sufficient stimulus exemplars (P. 356)."

The Individualized Curriculum Sequencing Model outlined by Holvoet, Mulligan, Schussler, Lacy, and Guess (1982) proposed that teachers should use varied materials, either within or across sessions, in order to maintain student interest and to enhance generalization. This recommendation has been supported by the findings of Horner and McDonald (1982) which demonstrated that general case instruction resulted in better generalization than did single-instance training with severely handicapped students. These authors used three different varieties of capacitors in the general case training format., and only one variety of capacitor in the single-instance training. They then tested generalization by having the student work with novel, untrained types of capacitors.

The investigation outlined below is similar to the study by Horner and McDonald (1982) in that a vocational task was studied, and multiple exemplars were used for training. It differs, however, in that the different exemplars (and probe items) all required slight variations in the response and were quite dissimilar in appearance. This study also does not attempt to compare multiple-exemplar to single-instance training, but seeks only to establish whether multiple-exemplar training that requires variations in a response results in generalization to untrained items, requiring similar responses.

Subjects

A male and female student were selected from a public school serving adolescents with special needs. Both students were selected from a classroom for severely handicapped individuals. The two students, Arthur and Kim, were 17 and 18 years of age, respectively. They were classified as severely retarded using the MS/CA ratio on tests such as WISC-R as well as on adaptive behavior measures such as the TARC Assessment Guide (Sailor & Mix, 1975).

Arthur could be described as a person in constant motion who stared off into space and flicked his fingers when not attending to the task given to him. Otherwise, he had good fine motor skill and enjoyed working on the tasks he was given.

Kim could be described as one who liked learning the tasks and knowing how to assemble them. She did not require very many verbal, demonstration, or physical cues, and seemed to learn quickly.

Setting

The study took place in a large well-lighted, vocational classroom of severely multiply handicapped students which was served by staff from a federally funded demonstration project. Each student was worked with individually. The student was seated at a long flat work bench next to the investigator. The reliability observers sat about five feet away, to the right of the student. Five other students and three staff were working nearby on leisure, vocational, and academic tasks unrelated to the research.

Tasks/Materials

There were six tasks used in this study. Three tasks served as the training items: assembling a metal faucet; screwing a metal hinge into a piece of wood; and screwing a metal handle into a piece of wood. Generalization probe items consisted of screwing a large bracket into a piece of wood, assembling a deadbolt lock on a piece of wood, and screwing two wooden handles into a piece of wood.

Training items. Each of the training tasks was divided into several sub-tasks (steps) that were taught in each session. On the metal faucet task, the student was required to place a knob on the faucet and then place one screw in a hole on top of the handle. The correct screwdriver was then chosen, and the screw tightened.

On the metal handle task, the student was required to place the handle in predrilled holes in the board, put the screw in by hand on the underside of the board, select the correct screwdriver, and then tighten the screws.

The metal hinge task required the student to place the hinge on a board, select the correct screwdriver, and place and tighten four screws.

Generalization items. Each of the generalization probe items was divided into several subtasks that were analyzed in each session. On the large bracket, the subject was to pick up the bracket and screw, place the bracket over the hole in the board, place the screw in the hole, select the correct screwdriver, and tighten the screw. The wood handle required that the subject first place two screws in predrilled holes, turn the wood over, and tighten the wood handles by hand. The student next needed to select the correct screw-driver, turn the board over again, and tighter the two screws.

On the deadbolt lock, the student was required to place the lock portion on the board, place four screws in the lock holes, select the correct screwdriver, tighten those four screws, place the clasp portion on the board, put in two more screws, and tighten them.

Measurement

The dependent variable in this study was measured on each of the six tasks. This was done for the training tasks by recording the level of prompting needed for a student to complete each step of a task correctly, using the least-prompts strategy described by Lent and McLean (1976). This strategy specifies four levels of prompting which are presented in sequence until the student makes an acceptable response. The levels consisted of: (+) = the subject performed independently; (V) = a verbal prompt is given; (D) = a demonstration prompt is given; (P) = the student is given physical assistance in completing the task. At the end of each task the number of steps the student had completed correctly at the independent level was determined. This total was then converted into a percent by dividing that total by the number of steps in the task.

In the generalization probe tasks, the investigator recorded whether or not the student completed any or all of the steps of the task independently. The total number of independent responses was then divided by the number of steps in that task.

Procedures

Baseline condition. In the baseline condition, the student sat next to the experimenter at the long work bench. The instructor gave the student the materials for one of the six tasks and three different types of screwdrivers were placed in front of the student. The order in which different tasks were given to the student were prearranged, and differed in consecutive sessions. The student then had an alloted amount of time to complete the task. For example, the student would be given the materials for the metal hinge and if he/she began working, he/she was given ninety seconds to complete the task. If the student did not pick up the materials and begin working in the first thirty seconds, then materials were removed and the next task was presented. If the student finished the item or quit working, the materials were removed after thirty seconds of "no work", or at the end of the ninety seconds, whichever occurred first. The responses were scored as correct (+), no response or incorrect response (-), and approximation (A). The instructor praised the student after the task was over by saying "that was a nice job" or "you worked pretty good on the (item)." Then the instructor would give the student a different item to assemble.

Multiple exemplar condition. In the multiple exemplar condition, the student was again seated next to the experimenter at the work bench. The instructor gave the student either a generalization probe item or a training item to assemble. When the student was given a training item, the instructor would say, "put this together." If the student did not respond correctly, the instructor would give a verbal, demonstration, or a physical prompt on the steps of the task. The responses were recorded on the data sheet as (+) for independent, (V) for verbal, (D) for demonstration, and (P) for physical prompt. Praise was provided specific to the step of the task the subject was engaged in. Arthur also received a pat on back, but Kim did not. When the student was given a generalization probe item, the instructor would hand the item to the student and say "put this together." The student's responses were recorded as correct or incorrect on each step of the task. The student was given the same amount of time to complete the probe item as in the baseline condition. The training and probe items were given in a different order every session. Praise was given at the end of the task; e.g., "nice job completing the (item)" or "I liked the way you worked on the (item)."

Experimental Design

Each task was analyzed in a multiple baseline format across subjects (Kazdin, 1973) to evaluate the effectiveness of using multiple exemplar training to enhance generalization. All tasks, whether training tasks or generalization tasks, were assessed for several sessions to determine the student's base level of performance without training. Multiple-exemplar training was then initiated across all three training tasks and the effect was assessed by observing the student's performance on the three generalization probe tasks. The onset of training was delayed more for Kim that it was for Arthur in this design.

Reliability

Reliability of the data was assessed for each student several times during the baseline and multiple-exemplar conditions. Reliability consisted of the instructor and an observer concurrently and independently recording the level of prompting on each step of the trained task, and the correctness of the response(s) on each step of probe tasks.

The instructor sat by the student at the work bench while the observer sat five feet away on another smaller work bench. At the end of the session, the instructor compared the two data sheets to see how many agreements there were between the two observers.

An agreement was scored if both the instructor and observer recorded the level of prompting in the same way on a step (e.g., both experimenter and observer put a (V) to indicate that a verbal prompt was needed to select the correct screwdriver on a task. Reliability was calculated by dividing the number of agreements by the number of agreements and disagreements, then multiplying by 100.

RESULTS

Reliability

The reliability measures for Arthur in baseline on the trained items were 90% on the metal faucet, and 100% on both the metal handle and metal hinge. For the generalization probe items in this conditions, the reliability measure was 100% on all three items. The reliability measures in the multiple exemplar condition for Arthur on the trained and generalization probe items was 100% on all items.

The reliability measures for Kim in the baseline condition for the trained items was 100%. Reliability measures in this condition, for the generalization probe items was 93% with a range of 86%-100% for the wood handle, with 100% on both the large bracket and deadbolt lock. The reliability measures in the multiple exemplar condition for the trained items was 100%. The reliability measures in this condition for the generalization probe items were 86% on the large bracket with a range of 75%-100%, and 100% for both wood handles and the deadbolt lock.

Training Items

Figures 1 through 3 include the data from the three training items for Arthur and Kim respectively. Figures 1-3 illustrate, graphically, the percent of correct responses made during each training session on the three training items (metal faucet, metal handle, and metal hinge). The training items are presented with Arthur's data on the top graph and Kim's on the bottom graph. These data are presented in this way for

Figure 1. The percent of independent correct responses made by the subjects on the metal faucet (a training task)

Metal Faucet **Training Data**

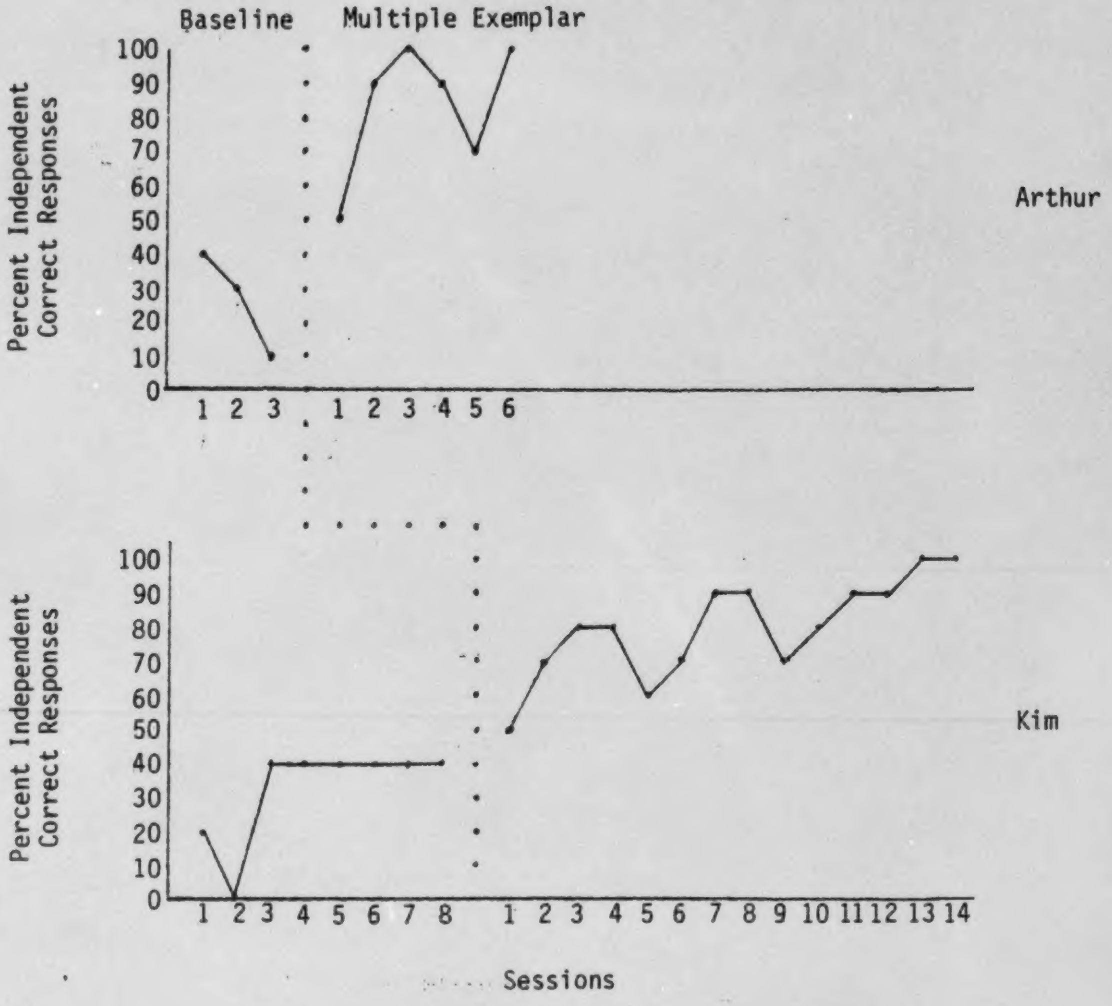


Figure 2. The percent of independent, correct responses made by the subjects on the metal handle (a training task).

Metal Handle Training Data

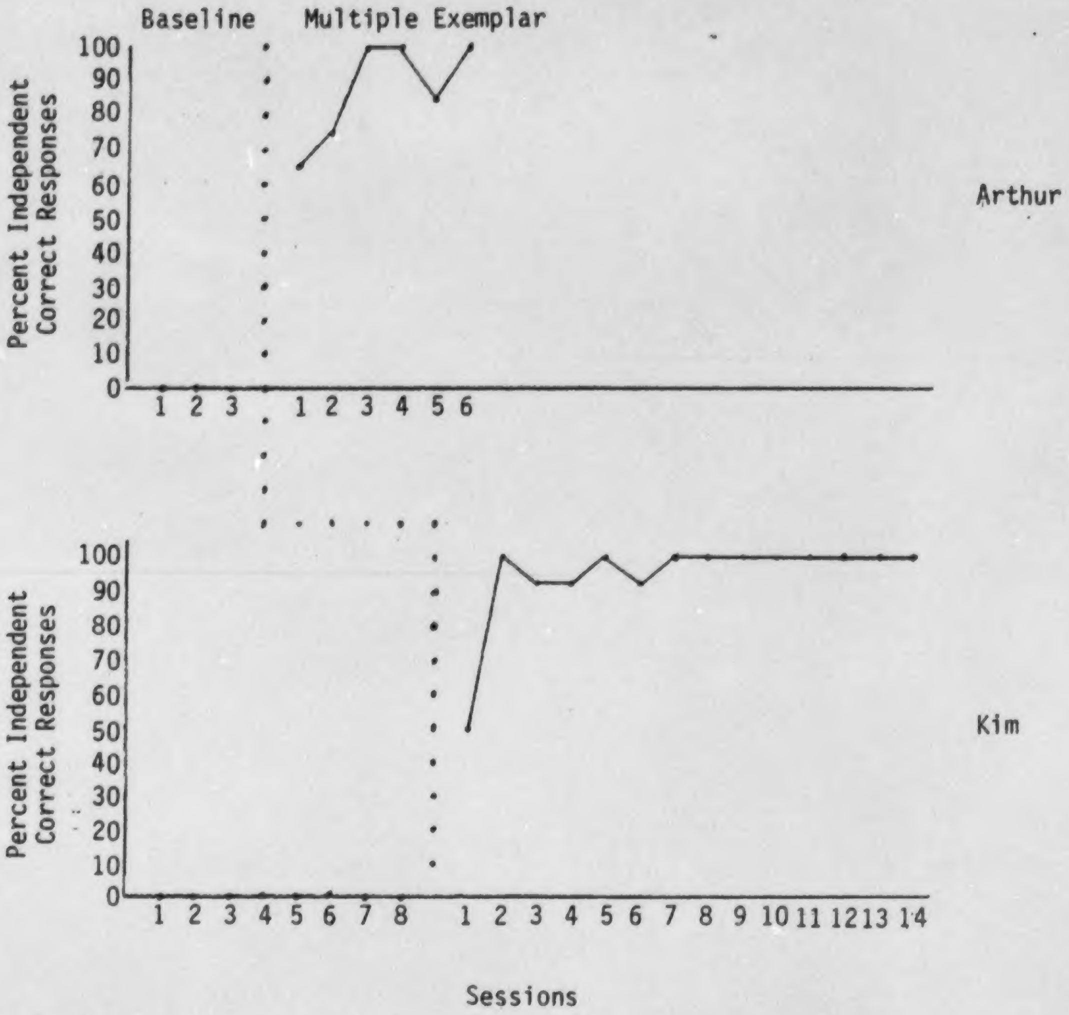
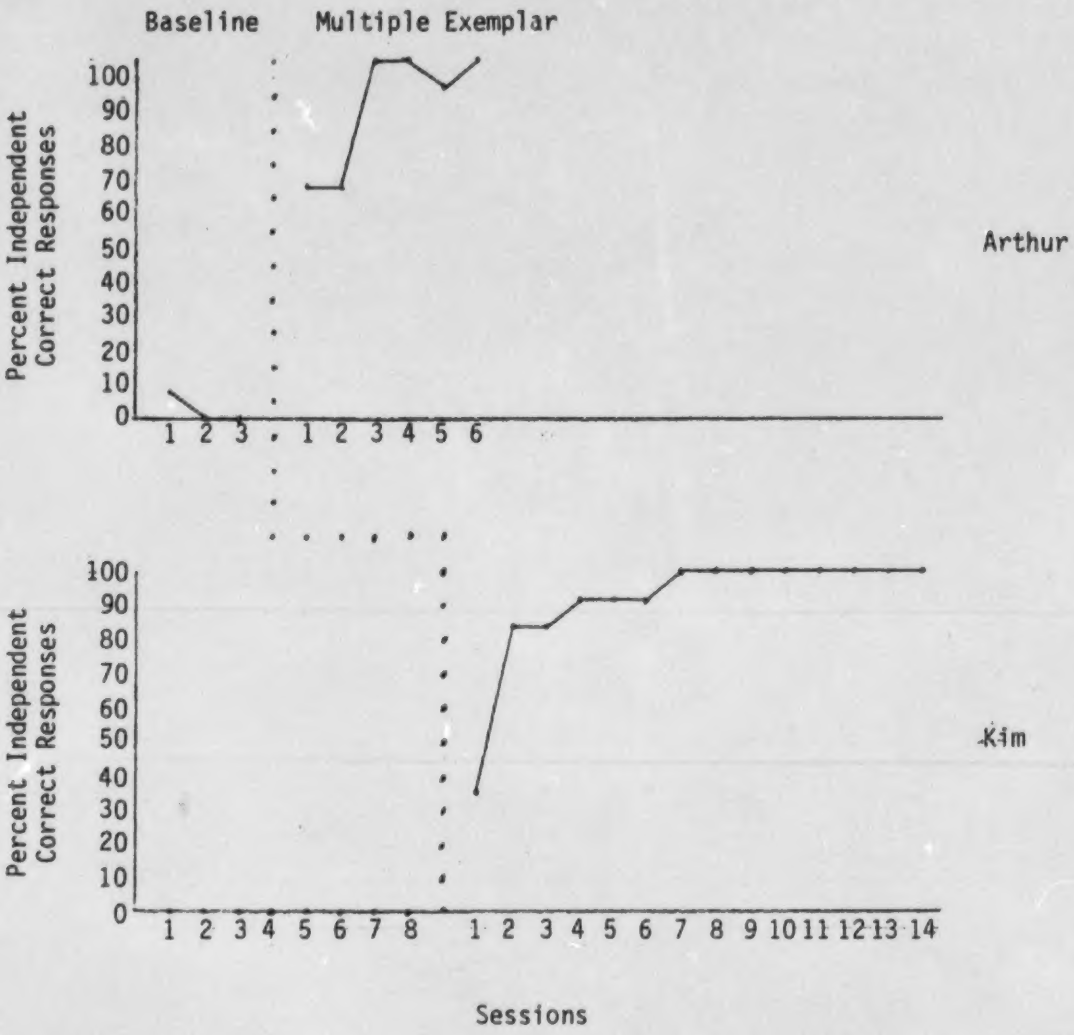


Figure 3. The percent of independent, correct responses made by the subjects on the hinge (a training task)

Metal Hinge Training Data



both baseline and mutliple exemplar conditions. These tasks were directly taught to the students during the multiple exemplar condition.

Item 1 (metal faucet). The baseline data for Arthur shows a descending trend with a mean score of 27%. In the multiple exemplar condition Arthur shows a generally accelerating trend. His mean score during training was 83%. Kim's data exhibit some acquisition during the third session of the baseline condition. The third session had 40% correct response and stayed at that level throughout baseline. Kim's baseline mean was 31%. Kim's data during multiple exemplar training showed a clearly ascending trend, and reached criterion on the eleventh session. Her mean score was 75%.

Item 2 (metal handle). The baseline data for Arthur revealed no acquisition on the metal handle (mean score was 0%). The multiple exemplar condition data for Arthur had an ascending trend with a mean score of 88%. Kim showed no acquisition in the baseline condition which gave her a mean score of 0%. Kim's data in the multiple exemplar condition also showed an ascending trend with criterion reached on the second session. Her mean score was 94%.

Item 3 (metal hinge). Arthur's baseline data initially showed 8% correct responses, then descended to zero for the next two sessions. Arthur's mean score in this condition was 3%. The multiple exemplar condition data on this item show that he reached criterion on the third training session. His mean score in this condition was 88%. Kim showed a zero performance on the metal hinge during the baseline condition for a mean score of 0%, over the eight sessions. The data for Kim in the multiple-exemplar condition showed a quick ascending trend, reaching criterion on the second training session. Her mean score in the multiple exemplar condition was 91%.

Generalization Probe Items

Figures 4 through 6 illustrate, graphically, the percent of correct responses made during each generalization probe session on the three probe items (large bracket, wood handle, and deadbolt lock). The generalization probe items are presented with Arthur's data on the upper graph and Kim's on the lower graph. These data are presented in this way in both conditions. No direct training occurred during the multiple exemplar condition for these items.

Item 4 (large bracket). During baseline, Arthur showed one day in which he made on correct response out of four possible. The other two sessions were at zero. His mean score was 8%. Arthur's multiple exemplar data showed that Arthur quickly showed generalization. Arthur's mean score was 92% in this condition. Kim showed no generalization during the eight sessions of baseline. She then exhibited a gradual learning of the task without training in the multiple exemplar condition. She had a mean score of 51%, but reached criteria and remained there on session eleven.

Item 5 (wood handles). Arthur's baseline data showed no generalization. Once he was in the multiple exemplar condition, he showed a rapid and high degree of generalization. He reached criterion four sessions in a row, had one session for poor performance, but returned to criterion on the next session. Arthur's mean generalization score was 93%. Kim's data show that after one day she acquired one correct response out of seven. That was maintained all through baseline condition. Her mean score was 13%. During this multiple exemplar condition, Kim showed a immediate increase in performance when multiple exemplar training was implmented, followed by a variable, but generally high performance. Her mean performance on this item was 94%.

Figure 4. The percent of independent, correct responses made by the subjects on the large bracket (a probe item)

Large Bracket **Probe Data**

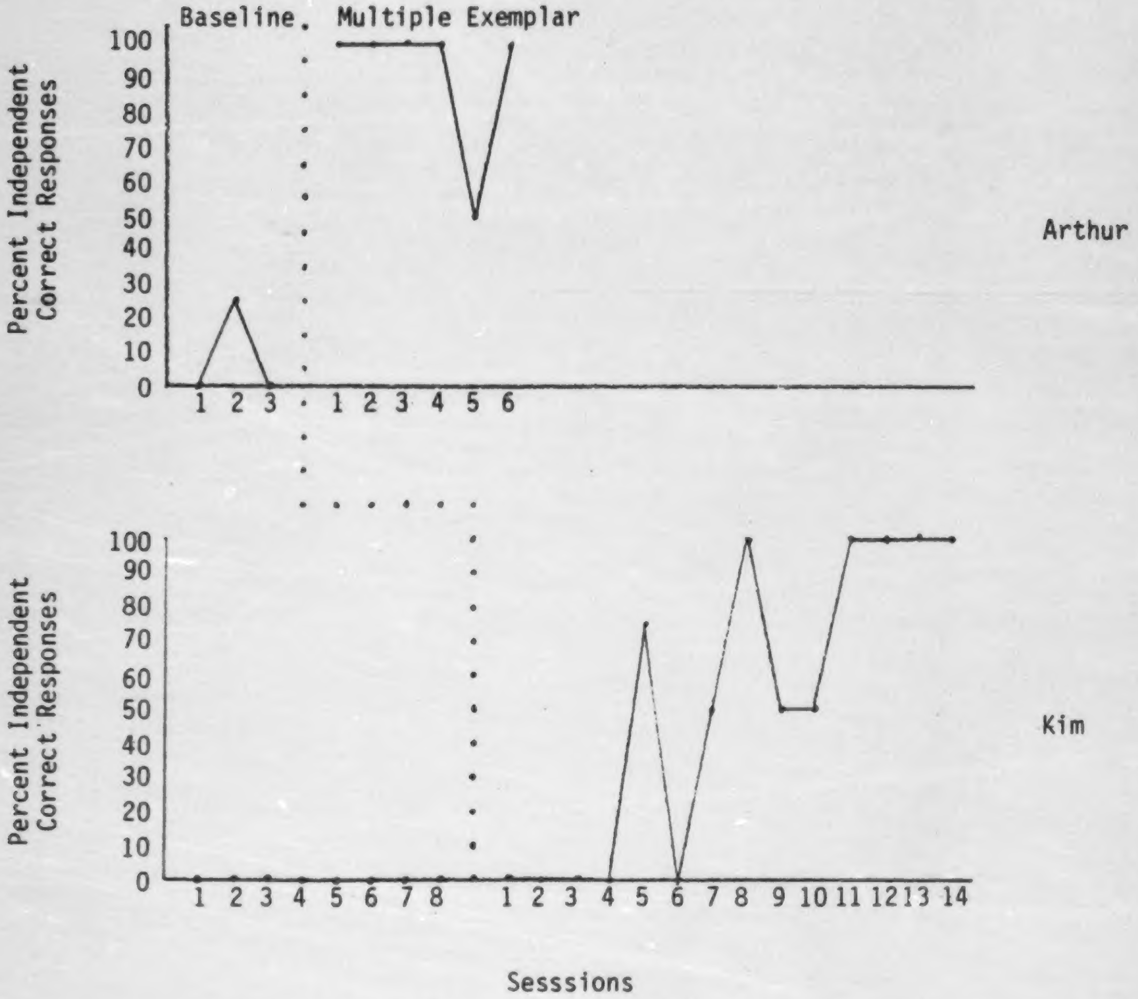


Figure 5. The percent of independent, correct responses made by the subjects on the wooden handles (a probe item)

Wood Handle **Probe Data**

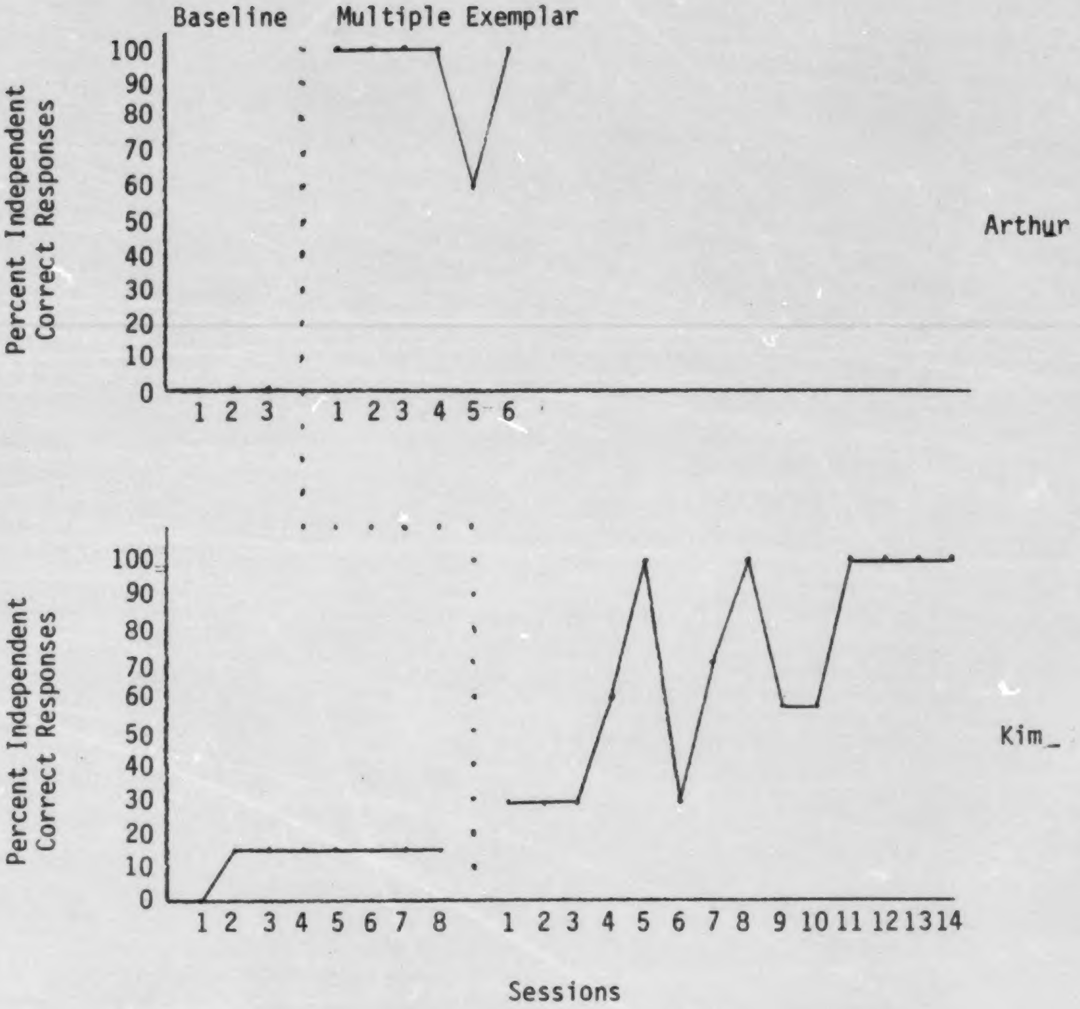
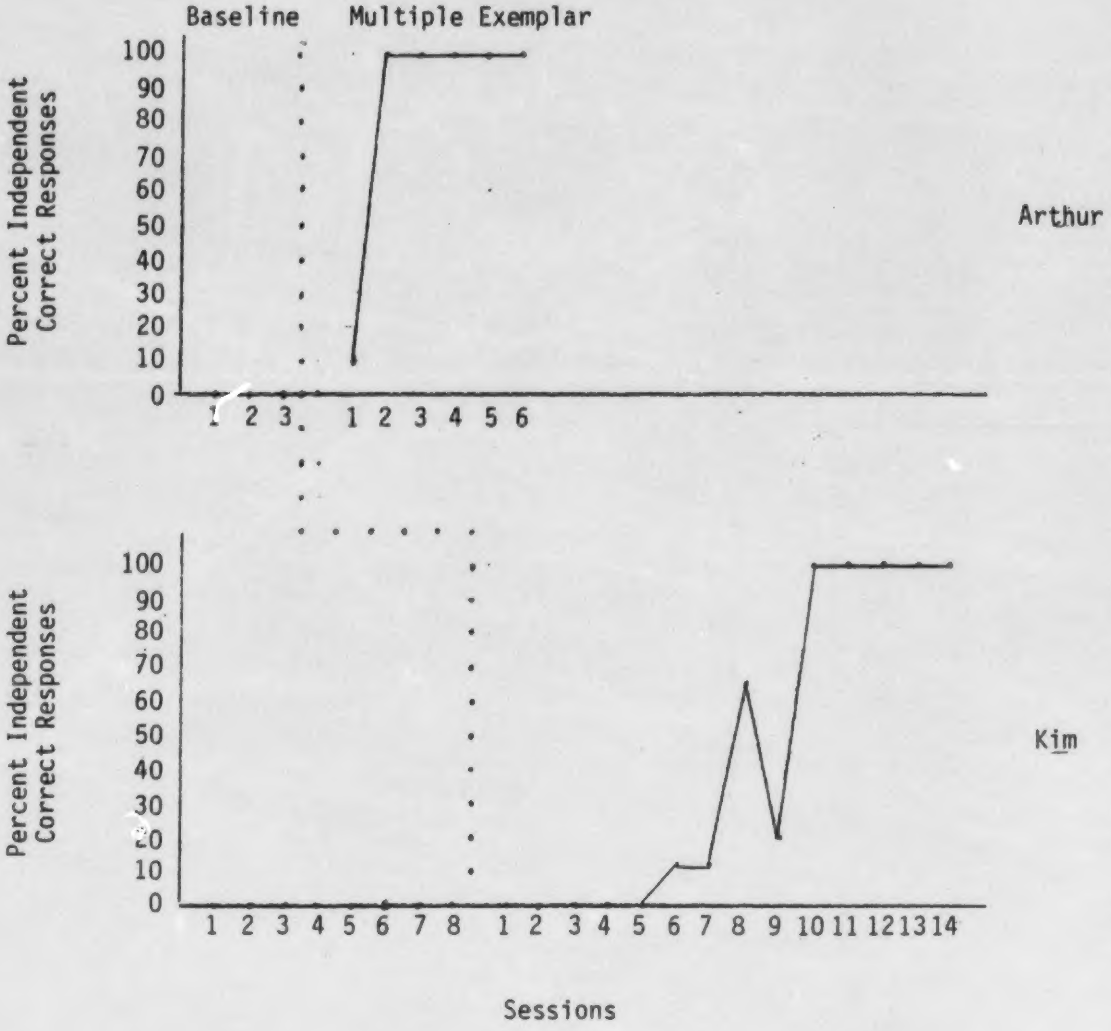


Figure 6. The percent of independent, correct responses made by the subject on the dead bolt lock (a probe item)

Dead Bolt Lock Probe Data



Item 6 (deadbolt lock). The baseline data displayed zero generalization for Arthur. Generalization performance of 100% on this item was reached in the second session of the multiple exemplar condition for Arthur. His mean generalization score in this condition was 83%. Kim's data showed zero generalization in baseline also. Kim's data in the multiple-exemplar indicated a five session period of no generalization, then she quickly reached criterion. Her mean score in this condition was 44%.

DISCUSSION

The multiple baseline across tasks design allowed adequate experimental control to affirm: (a) that there was an increase in acquisition of the probe items for Arthur as soon as training began in the multiple exemplar condition; (b) that there was a gradual increase in the acquisition of the probe items for Kim on the multiple exemplar condition, and (c) the training of multiple exemplars was functionally related to the increase of correct responses on the probe items.

These data, from Arthur and Kim, support the hypothesis that learning to perform a task with a variety of items increases generalization across similar untrained tasks. Thus, it would seem that it would be good educational practice to use a variety of materials to teach severely handicapped students who traditionally have difficulty in generalizing. The ability to generalize might enable these students to be more independent in non-school settings.

These data extend the findings of Horner and McDonald (1982), showing that teaching across physically dissimilar materials that require the same topographical response is also effective in obtaining response generalization.

These results, however, must be cautiously interpreted due to certain design limitations. First of all, a critical assumption is that performance on the three generalization probe items is representative of what the student's performance would be on all other items which require screwdriver assembly. It is possible that a wider range of probe items might have pointed out generalization deficiencies that were not apparent in this study.

A second area of potential difficulty is that instructor bias might have influenced the results. One instructor conducted the baseline and multiple exemplar training with both students. This same instructor also conducted the generalization probe sessions, in order not to add the variable of generalization across persons. The instructor was aware of the purpose of the study and the experimental hypothesis. It is certainly possible that uncontrolled, inadvertant changes in his behavior might have affected student behavior. While the promptness and extent of improvement following the introduction of multiple exemplar training makes this an unlikely possibility, the opportunity for bias forces some reservations in interpreting the results.

A third limitation is that both this study and the Horner and McDonald (1982) study focused on tasks found in the vocational domain. It is possible that the obtained generalization is unique to vocational tasks or to tasks that require fine motor manipulation. The premise that multiple exemplar training improves generalization needs to be tested across self-help tasks (e.g., trying different types of shoes), independent living tasks (e.g., using different types of washing machines), and leisure tasks (e.g., shuffling different sizes and thicknesses of cards) before it is applied across these domains in the classroom.

Lastly, it should be emphasized that the results of this study only apply to generalization across tasks. It is obvious that a student must also learn to generalize across settings and across persons if the skill is to be optimum value. It is possible, but not certain, that using multiple training sites or multiple teachers would result in these types of generalized response. Such an assumption needs to be investigated. Also it would be fruitful to determine whether a training situation where all three aspects (task, setting, and teacher) were systematically varied would result in efficient generalization or a very confused student.

In spite of these limitations, this study supports the idea that using a variety of objects to teach a vocational skill is an efficient way to encourage generalization across untrained objects.

REFERENCES

Brown, L., & York, R. Developing programs for severely handicapped students: Teacher training and classroom instruction. Focus on Exceptional Children, 1974, 6(2), 24-36.

Drabman, R. S., Hammer, D., & Rosenbaum, M. S. Assessing generalization in behavior modification with children. The generalization map. Behavioral Assessment, 1979, 1, 203-219.

Holvoet, J., Mulligan, M., Schussler, N., Lacy, L., & Guess, D. The KICS model: Sequencing learning experiences for severely handicapped children and youth. Lawrence, KS: University of Kansas, Department of Special Education, 1982.

Horner, R. H., & McDonald, R. S. Comparison of single instance and general case instruction in teaching a generalized vocational skill. Journal of the Association for the Severely Handicapped, 1982, 7(3), 7-20.

Johnson, S., & Bolstad, O. Methodological issues in naturalistic observation: Some problems and solutions for continued research. In L. Hammerlynck, L. Handy, & E. Mash (Eds.), Behavioral Change: Methodology, concepts and practice. Champagne, IL: Nescarch Press, 1973.

Kazdin, A. E. Methodological and assessment considerations in evaluating reinforcement programs in applied settings. Journal of Applied Behavior Analysis, 1973, 6, 517-531.

Lent, J. R., & McLean, B. M. The trainable retarded: The technology of teaching. In N. G. Haring & R. L. Schiefelbusch (Eds.), Teaching Special Children. NY: McGraw-Hill, 1976.

Sailor, W., & Mix, B. J. The TARC assessment system. Lawrence, KS: H & H Enterprises, 1975.

Stokes, T. F., & Baer, D. M. An implicit technology of generalization. Journal of Applied Behavior Analysis, 1977, 10, 349-367.

Wehaman, P., Abremson, M., & Norman, C. Transfer of training in behavior modification programs: An evaluative review. Journal of Special Education, 1977, 11, 212-231.

Evaluation of the Effects of Group Instruction

The Individualized Curriculum Sequencing Model porposed that structured group instruction should be an integral part of an SMH curriculum. The studies outlined in this section looked at the effects of structuring interactions between two students being taught by the same teacher. Two of the studies looked at incidental learning and asked whether placing students in a group learning situation results in them learning one another's programs (incidental learning). Of particular interest was whether structuring an interaction (such as passing materials or edible reinforcers) during the group learning situation improved incidental learning.

The third study investigated whether using a group format which mandated some interaction between students (i.e., an assembly-line) was superior to a format where students worked alone on vocational tasks. This study focused on the impact of such group arrangements on students who had a history of having difficulty staying on task in a workshop situation.

For the most part, the studies supported the use of groups for teaching students with severe multiple needs.

The Effect of Systematic Peer Interaction on the Incidental Learning of Two Severely Handicapped Students[1,2]

Fredda Brown, Ph.D.
Division of Professional Education
State University of New York at Binghmaton

Jennifer Holvoet, Ph.D.
Department of Special Education
University of Kansas

1982

---

[1] This paper is reprinted with permission from the Journal of the Association for the Severely Handicapped, 1982, 7(4), 19-28.

[2] Portions of this paper were prepared as a part of the work pursuant to the "Kansas Individualized Curriculum Sequencing Project" funded as a Handicapped Children's Model Project, U.S. Office of Educaiton (G00-800-1721) from 9-1-80 to 9-1-83.

Abstract

This study investigated the effects of systematically including peer interaction trials during a group training session for two severely handicapped adolescent students. Both students in the dyad were directly taught different three-component sequences of skills. Using a multiple baseline design, incidental learning was assessed by measuring each student's performance of the other students' skill sequence. This study also explored whether the placement of the peer interaction in the sequence (i.e., whether it occurred after the first, second, or third component of the sequence) affected incidental learning. Results indicated that for one student incidental learning occurred with only a minimal amount of structured interaction between the students in the dyad. Results further indicated for this student a direct relationship between the placement of interaction in the skill sequence and the acquisition of the specific skills. For two of the three skills, the student showed incidental learning of the skill presented just prior to the peer interaction trial.

The Effect of Systematic Peer Interaction on the Incidental Learning of Two Severely Handicapped Students

Group instruction for the retarded student has, in the past, been viewed as an unfortunate alternative to 1:1 training, to be used only when the student-teacher ratio is too high. More recently, however, group instruction has become a meaningful and purposeful part of individualized education programs. Brown, Holvoet, Guess and Mulligan (1980) maintain that group instruction must be structured, data-based and derived from the preassessed needs of each student. The authors delineate the goals of instruction in a group setting as three-fold: maintaining skills that were acquired in other settings, improving already learned skills (e.g., increasing rate of performance, decreasing dependence on continuous reinforcement and cues) and acquiring new skills.

Many researchers have stressed the importance of group instruction, for handicapped individuals. Generally, this has been for three reasons: normalization (Barrett, 1979; Brown, Nietupski, & Hamre-Nietupski, 1976), efficiency (Alberto, Jobes, Sizemore & Doran, 1980; Favell, Favell & McGimsey, 1978; Kohl, Wilcox & Karlan, 1978; Storm & Willis, 1978), and opportunity for incidental learning (Brown, et al., 1976; Orelove, 1982). Orelove (1982) described incidental learning as situations where one student in a group learns materials presented to one or more other group members; thus the student who learns incidentally demonstrates knowledge about some aspect of a task that was not explicitly demanded of him/her. Such opportunity for incidental learning through the observation of peers, however, is very limited in the traditional special education class for severely handicapped students where intensive 1:1 instruction is the typical mode. The use of group or peer-to-peer instructional techniques should increase this opportunity. It cannot be assumed, however, that because the opportunity is provided to observe another, that in fact one student will attend to the other; a person cannot learn much by observation if the individual does not attend to the important features of what is being observed (Bandura, 1973). Many researchers have demonstrated that just exposing the handicapped individual to various models was not sufficient to change behavior (Guralnick, 1976; Snyder, Spolloni & Cooke, 1977; Sullivan, Note 1; Yelenik, Note 2).

It seems however that some type of environmental manipulation or systematic structure may be necessary to facilitate learning from others in the environment. Ruggles and LeBlanc (Note 3) found that even with nonhandicapped children systematic interaction between students resulted in more observational learning than in groups where such interaction was encouraged but not systematically elicited. Mithaug and Wolfe (1976) used a form of structured peer interaction where correct responding on a given trial was dependent on information that was available only from a peer. They found that task interdependence was sufficient to increase verbalizations with three of the four students; when a verbal contingency was added to the interaction, all four of the students' verbalizations increased. Brown et al. (1980) describe a model of group instruction that formally programs interaction between the students in a group or dyad. The Intersequential group structures an interaction between two students by having the students be antecedents or consequences to each other's behavior.

The present research compared the effects on incidental learning of a group that employed a structured student interaction to a group that did not formally structure student interaction. Having the students reinforce each other, using tangible reinforcers, was chosen as the structured interaction. Incidental learning was measured by each student's performance on a sequence of three skills that were directly taught to the other student in the dyad.

Three specific questions were addressed in the study:

1- Will one student learn the other student's tasks by just being physically present in the same dyad; i.e., no structured peer interaction?

2- Will one student learn the other student's task when prompted to reinforce the other student's correct responses; i.e., when there is a structured interaction between the students?

3- At what point in the skill sequence is the interaction (i.e., the prompt to reinforce) most effective in producing learning of the other student's task?

METHOLOGY

Subjects

Subjects were two adolescent male students from two different self-contained classrooms for severely retarded individuals. Dave was 17.5 years old and had resided in an institution for most of his life. The Merrill Palmer Scale of Mental Tests (Stutsman, 1931) and the

Vineland Social Maturity Scale (Doll, 1965) indicated that Dave was functioning in the profound range of retardation. On the TARC Assessment Inventory (Sailor & Mix, 1975), administered by the teacher, Dave scored at the 40th percentile, scoring the lowest in the expressive communication and pre-academic skills. This student showed fair receptive language, good fine motor skills and the ability to imitiate an adult's gross motor actions when specifically instructed.

Bill was 12.5 years old and suffered from grand mal status seizures that were only partially controlled with high dosages of medication. This student lived at home. On the TARC Assessment Inventory, Bill scored at the 30th percentile. On the Callier-Azusa Scale (Stillman, 1978), he performed at or below the 18 month level. He showed marked deficits in both receptive and expressive communication and was non-imitative. In the classroom situation this student showed no gestural or other pre-linguistic skills unless physically guided. He also did not follow verbal instructions and frequently spit at, hit or kicked peers and adults.

Both students enjoyed candy, crackers, pretzels, and cookies. The use of these edibles in other classroom programs demonstrated their reinforcing value to the students. These edibles were chosen to consequate correct behaviors during the training sessions.

Setting

Training took place in a self-contained classroom located in a public school building. Training was implemented during regular classoom hours in the group instruction area of the classroom. The group instruction area contained two desks and chairs, positioned adjacently for the two students, and one chair facing these for the trainer.

Instructional Programs

A three-part sequence of skills (i.e., a sequence composed of three different programs) was developed for each student. Objectives were chosen that were age-appropriate, attainable and related to the IEP goals for both students.

The terminal objectives in each sequence that were selected for the study were:

Bill

1. Opening an envelope in which there is an advertisement
2. Manually signing "advertisement"
3. Tacking the advertisement on a corkboard

Dave

4. Manually signing "Picture" when shown a photograph
5. Finding an empty slot in a photograph album page
6. Sliding the photograph into the slot.

Task analyses were then written for each of the tasks and the students were pretested to determine on which step of each program they should receive training. The individual steps that were selected for the study were:

Bill

1. Given a manilla envelope with the clasps pulled into a vertical position, the student will lift the flap of the envelope and reach into the envelope.

2. The student is positioned by the trainer so that his left hand is fisted and against his mouth, and his right hand is fisted and approximately six inches in front of his left hand. The student is required to bring his right hand in contact with his left hand.
3. Given a tack pushed half-way into the corkboard, the student will push the tack the rest of the way.

Dave

4. With the left hand the student will form the sign for the letter "c" adjacent to his left eye; the sign will then be brought down to touch the palm of the right hand.
5. Given a photograph album page of 5 empty slots and one slot filled with a photograph, the student will point to an empty slot. The number of empty slots will decrease as the trials proceed in the session.
6. Given a photograph halfway into the slot, the student will slide the remainder of the photograph into the slot.

Instruction Sessions

Teaching component. Each day Bill and Dave were instructed on their respective sequence of three skills. The teaching alternated between the students, such that one student would be asked to do his sequence of three behaviors, then the other student would be asked to do his three behaviors. This alternation of the sequences was repeated ten times, thus, there were five trials given for every program in the sequence. With each program, modeling of the correct response and physical guidance were used to correct student errors. Data were taken on each of the trials and edible reinforcers were given for correct responses, even if these responses had to be prompted. These reinforcers were delivered either by the experimenter or the other student as prescribed in the subsequent Experimental Design section. During training each student had two roles: he was an instructed student when the teacher was directly teaching him, and he had the opportunity to be an observer when it was his peer's turn for direct instruction.

Probes. When the ten trials of the Teaching Component had ended, the students began the Probe Component where the degree of incidental learning was assessed. In this component each student was given two opportunities to do the tasks that he had only observed during the teaching component. Bill was given a chance to do Dave's three programs once, then Dave was given a chance to do Bill's three programs. This alternation was then repeated a second time, for a total of two probes for each student. During the Probe component, verbal praise was given for correct responses, however, no praise, reinforcement or physical assistance was given for approximations of the correct response or incorrect responses.

Two types of probes were employed to assess the incidental learning of the two students: 1) The Individual Step probe assessed acquisition of that step of the instructional program on which the other member of the dyad was currently being instructed. The Individual Step probe allowed the student to demonstrate those skills that he had the opportunity to directly observe and 2) The Terminal Objective probe examined the acquisition of the terminal or final step of the instructional program. This type of probe allowed the student the opportunity to demonstrate mastery of the terminal instructional objective. The two probes were alternated daily. The only exception to the two-probe format was the "find the empty slot" task where the instructional program contained only one step so that the Individual Step probe and the Terminal Objective probe were identical.

Experimental Design

A multiple baseline design across behaviors (i.e., across the three instructional programs) was used for each member of the dyad. There was, therefore, a total of two multiple baselines used to assess incidental learning. The dependent variable, incidental learning,was measured during the Probe component of the sessions. In the multiple baseline design across behaviors, a period of no intervention (baseline) is followed by a period of intervention on a series of tasks or behaviors. The intervention is sequentially introduced to each task at a different point in time. In this study, the baseline was that period of time when there was no structured peer interaction. This was followed by a period in which the experimental intervention (structured peer interaction) was applied. The introduction of the experimental variable was lagged across time for the tasks.

In this study, the independent variable was structured peer interaction that occurred in the Training Component of the sessions. Structured peer interaction was defined as requiring the observing student to pass an edible to the student being directly instructed at the conclusion of each trial. The baseline (no structured peer interaction) was defined as the teacher giving all cues, prompts, and edibles. The dependent variable (measured during the Probe component of the session) was the degree of incidental learning demonstrated by each student (i.e., the accuracy with which a student who had only observed another student's program could do the behaviors). A further description of the experimental conditions follows:

Baseline (No structured peer interaction). This condition offered no structured interaction between the two students. During this condition the teacher alternated instruction between the two students. The materials, cues, prompts and edible reinforcers were all presented by the teacher.

Experimental treatment conditions (Structured peer interaction). During this condition the teacher alternated instruction between the two students. The materials, cues and prompts were presented by the teacher. The edible consequences for one of the skills in the sequence, however, were delivered by the observing student. In order to accomplish this, the trainer gave whatever verbal or physical assistance was necessary to get the observing student to deliver the reinforcer. The teacher still delivered the edibles for the other two skills in the sequence.

For example, if Dave was to be giving the edible to Bill for Skill 1 in the sequence (open envelope), the following would occur:

a) The teacher would implement one trial of Skill 1 (open envelope) for Bill, using the training procedures outlined in the instructional procedures;

b) The trainer would then put the reinforcer to be delivered (e.g., candy, cookie, cracker) on Dave's desk and say;

c) "Look, Bill just finished" or "Bill opened the envelope" (depending on how Bill performed the task)...'Why don't you give him the cookie/cracker/candy?"

d) If the student did not begin the interaction following the verbal prompt, the trainer physically assisted the student to deliver the edible.

The teacher continued to deliver the edibles for Skill 2 and Skill 3, as described in the baseline condition. Then, in the Probe trials Dave would be assessed on his ability to open the envelop (as well as perform any other of Bill's tasks) to see if there was any change from the No Structured Peer Interaction condition; i.e., if incidental elarning had occurred.

There was a total of three Structured Peer Interaction conditions for each student; a peer interaction was arranged for each instructional program in the skill sequences (i.e., multiple baseline across behaviors). The teaching component during each of the Structured Peer Interaction conditions can be represented as follows:

Structured Peer Interaction I

Bill		Dave
Skill 1		
Skill 2		
Skill 3		
	←	reinforce
		Skill 4
reinforce	→	
		Skill 5
		Skill 6

Structured Peer Interaction II

Bill		Dave
Skill 1		
Skill 2		
	←	reinforce
		Skill 4
		Skill 5
reinforce	→	
		Skill 6

Structured Peer Interaction III

Bill		Dave
Skill 1		
	←	reinforce
Skill 2		
Skill 3		
		Skill 4
		Skill 5
		Skill 6
reinforce	→	

Reliability. Interobserver reliability was taken on both the Instructional program and the dependent variable (i.e., incidental learning). Reliability was assessed on the Individual Step probe at least once in each Intersequential condition. A total of seven reliability checks were made. Agreement scores were calculated by dividing the number of agreements by the number of agreements plus disagreements and multiplying by 100. Agreement was defined as both the teacher and the observer scoring the same code for correct responses, approximations and incorrect responses (+, ⊕, or -).

Results

Interobserver reliability was taken separately for acquisition of the instructional program and for incidental learning. Reliability data on the instructional programs ranged from 87%-100% across the seven observations. Interobserver reliability on incidental learning resulted in three scores of 83% and four scores of 100%. The disagreements during these sessions were concerned with the interpretations of "approximation" responses.

Three specific questions were addressed in this study: 1) Will one student learn the other student's tasks by just being physically present in the same dyad (i.e., in the baseline condition); 2) Will one student learn the other student's tasks when an interaction is structured between the two students; and 3) At what part of the skill sequence is the interaction most effective in producing learning of the other student's tasks?

The first question was answered by examining the Baseline (No Structured Peer Interaction) condition for each student during the first three days of training. Although the multiple baseline extends the length of the baseline condition across varying amounts of time for each behavior, there are only three days of this baseline that do not have any interaction conditions in effect. If incidental learning occurred at this time then it could be assumed that the physical proximity that occurs in a group was responsible for the learning. Figures 1-4 indicate that except for one "approximation" made during this time by Dave; there was no evidence of any acquisition of the other students' task.

The second and third questions were addressed by implementing the Structured Peer Interaction conditions. If there was acquisition across all tasks for each student with the initial structured interaction condition, then a generalized effect of the interaction could be assumed. However, if acquisition consistently occurred only on the task that was involved in the interaction, then incidental learning could be assumed to be directly related to the placement of the interaction within the skill sequence.

Terminal Behavior Probe (Probe Component)

Figure 1 represents Dave's incidental learning of Bill's task when probed on acquisition of the terminal objectives. Implementation of the first two structured interaction conditions (structured interaction followed "tack ad" and "sign ad") resulted in a rapid increase in "approximations" to the terminal objective. This increase in "approximations" rather than "corrects" was expected because the student was not directly observing the terminal objective, but an earlier step in the instructural program. Implementation of the final Structured Interaction Condition (structured interaction followed "Open Envelope") did not result in any increases in either approximated or correct responses. Withdrawal of the conditions did not seem to influence the acquired responses; that is, once incidental learning of a skill was demonstrated, it was maintained even when the structured interaction condition was no longer in effect.

Figure 2 represents Bill's incidental learning of Dave's tasks when probed in acquisition of the terminal objectives. Following implementation of the first Structured Interaction condition there seems to be a generalized effect (although inconsistent) on correct responses on the "finding slot" task. At this time there is also a single approximated

Figure 1. Dave's incidental learning of the terminal objectives of the three tasks that Bill was directly instructed on. Incidental learning was assessed during the Probe component of the session.

Dave: INCIDENTAL LEARNING-
TERMINAL OBJECTIVE PROBE

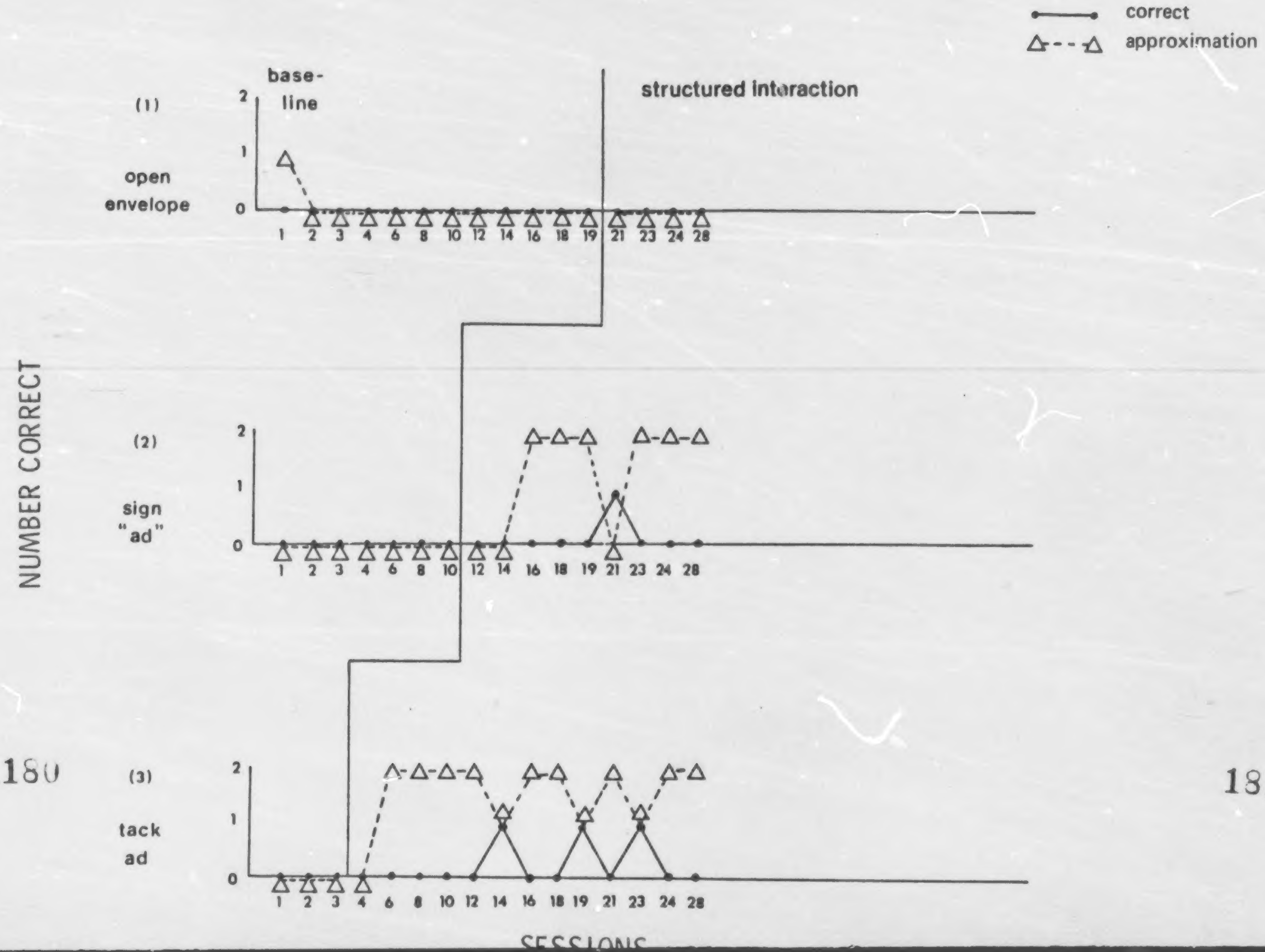
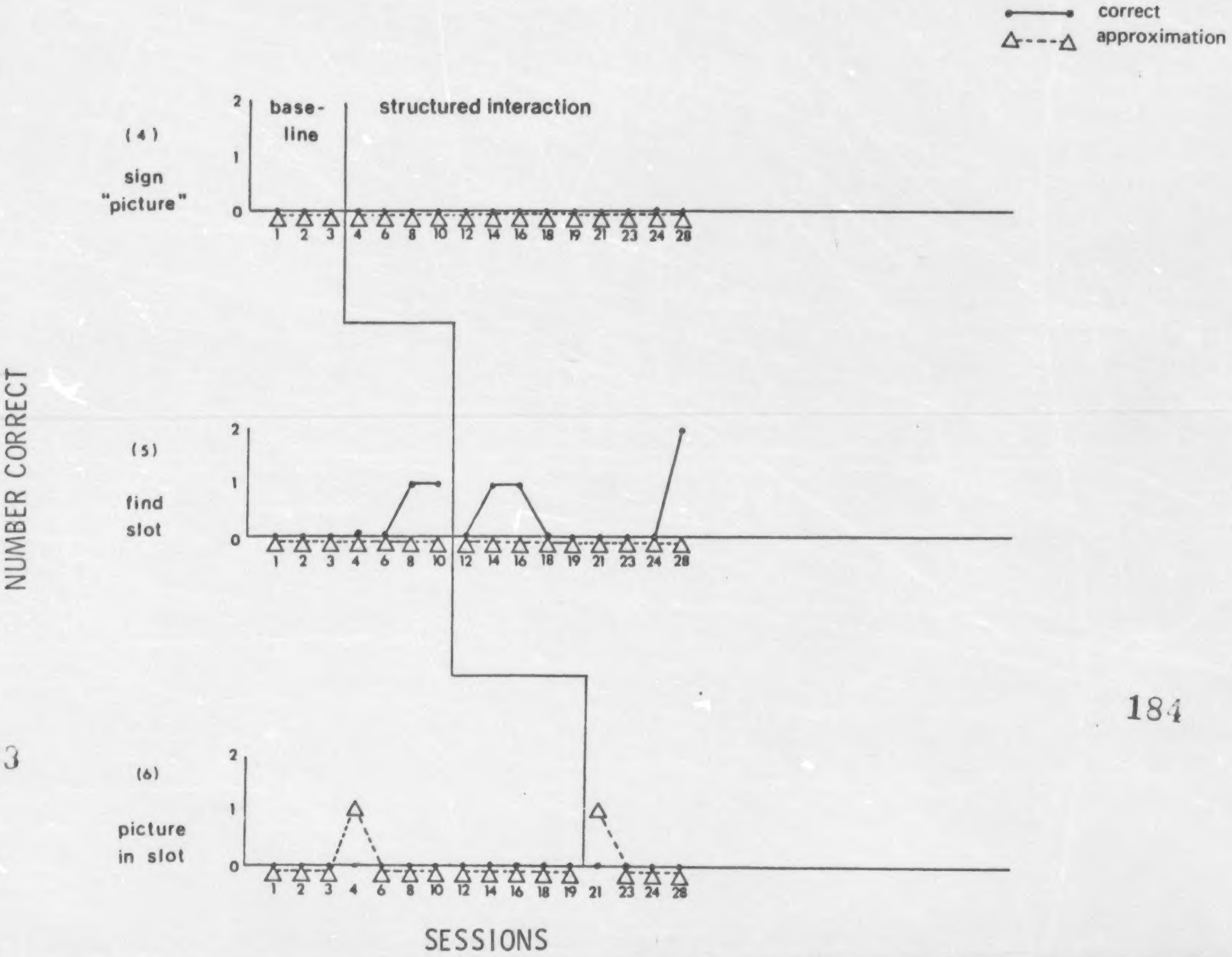


Figure 2. Bill's incidental learning of the terminal objective of the three tasks that Dave was directly instructed on. Incidental learning was assessed during the Probe component of the session.

Bill: INCIDENTAL LEARNING-
TERMINAL OBJECTIVE PROBE



response of "Picture in slot." Although these data are not strong, the placement of these occurrences is worth noting.

INSERT FIGURE 1 AND 2
ABOUT HERE

Individual Step Probe (Probe Component)

The Individual Step Probe assessed the students more directly on what they were observing. This probe was therefore expected to result in more correct responses than approximated responses. Figure 3 shows Dave's incidental learning of Bill's tasks using this probe. Implementation of the first Structured Interaction (tack ad) resulted in immediate correct responding. The second Structured Interaction (sign ad) resulted in an immediate, although not lasting correct response which then fluctuated between approximations and correct responses. The final Structured Interaction (open envelope) indicates only a single correct response following implementation of that condition. Similar to the Terminal Objective probe, withdrawal of the conditions did not seem to influence the acquired responses.

Figure 4 represents Bill's incidental learning of Dave's tasks when probed on the individual steps. Implementation of the first condition showed no effect on either approximated or correct responses of any of the three tasks. The second Structured Interaction had an immediate, but not lasting effect on the "find slot" task. The third Structured Interaction had an immediate effect on "picture in slot"; responding in this condition changed from one correct response to approximated responses. Although Bill's data were not very dramatic it is interesting that what little responding there was, did occur directly following implementation of the interactions.

INSERT FIGURES 3 AND 4
ABOUT HERE

In summary, these data show significant effects on two of the three Intersequential conditions for Dave both on the Terminal Behavior probe and on the Individual Step probe. Whereas the Terminal Behavior probe resulted in many approximated responses, the Individual Step probe resulted in more correct responding. Bill's data, for both types of probe, do not show very significant results; however, the few responses that were might indicate a possible relationship to the Structured Interaction conditions.

Discussion

The present study confirms the growing importance of group instruction with severely handicapped students. Most broadly the results of the study indicated that for one student, incidental learning did occur in a group of two students with only a minimal amount of structured interaction. The effects of the introduction of the Structured Interaction conditions at different points in Dave's skill sequence suggest a possible relationship between the structured interaction and incidental learning. One possible explanation for this might be related to Bandura's (1973) stress on attentional processes, that is, the interaction might serve to cue the student to attend to just that part of the student's skill sequence. It is interesting to also note that when Dave incidentally learned a skill, that skill was maintained, even when the interaction was withdrawn. Perhaps this is so because during the probe sessions the student was being praised for correct responses.

Figure 3. Dave's incidental learning of the three tasks that he observed Bill being directly instructed on. Incidental learning was assessed during the Probe component of the session.

Dave: INCIDENTAL LEARNING- INDIVIDUAL STEP PROBE

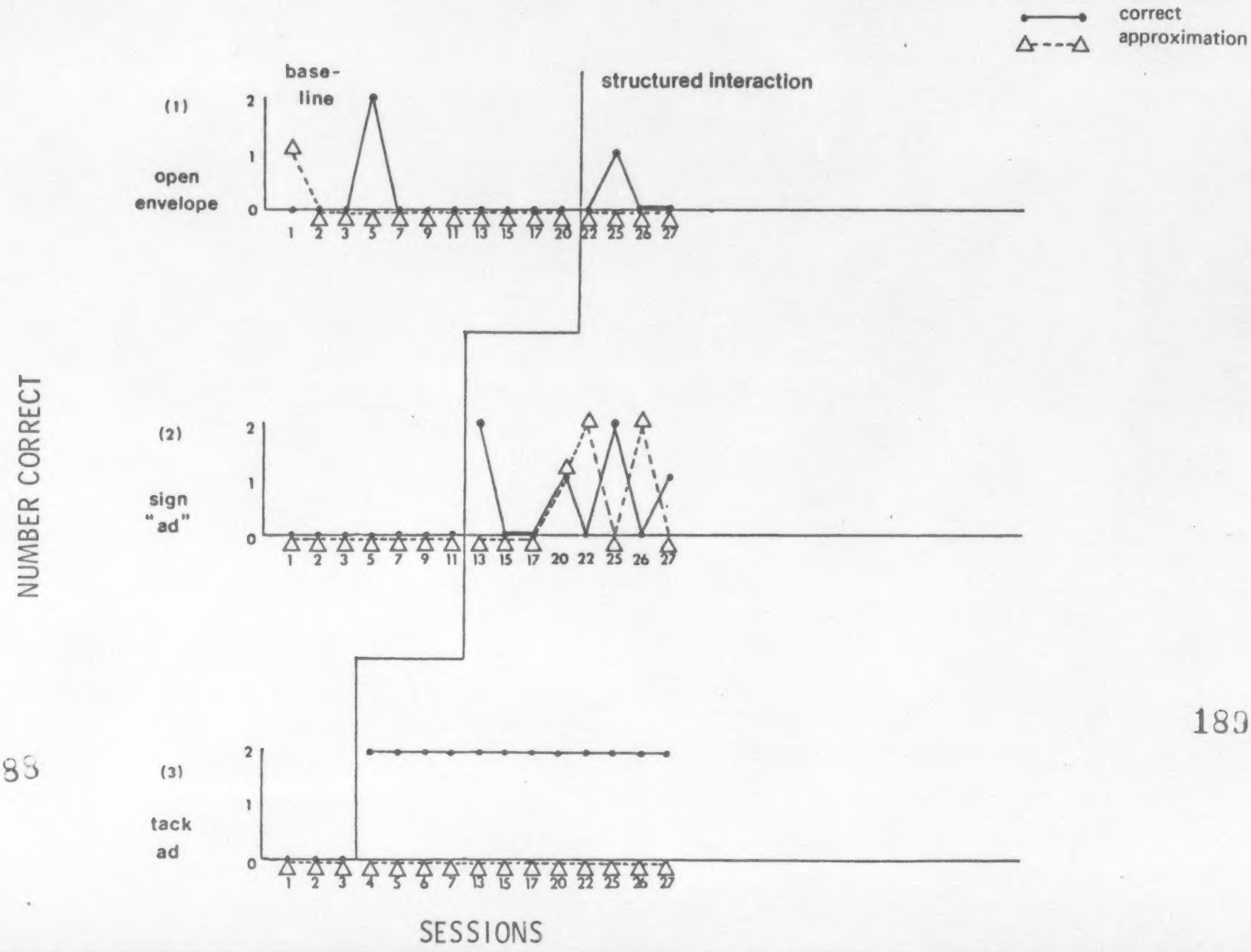
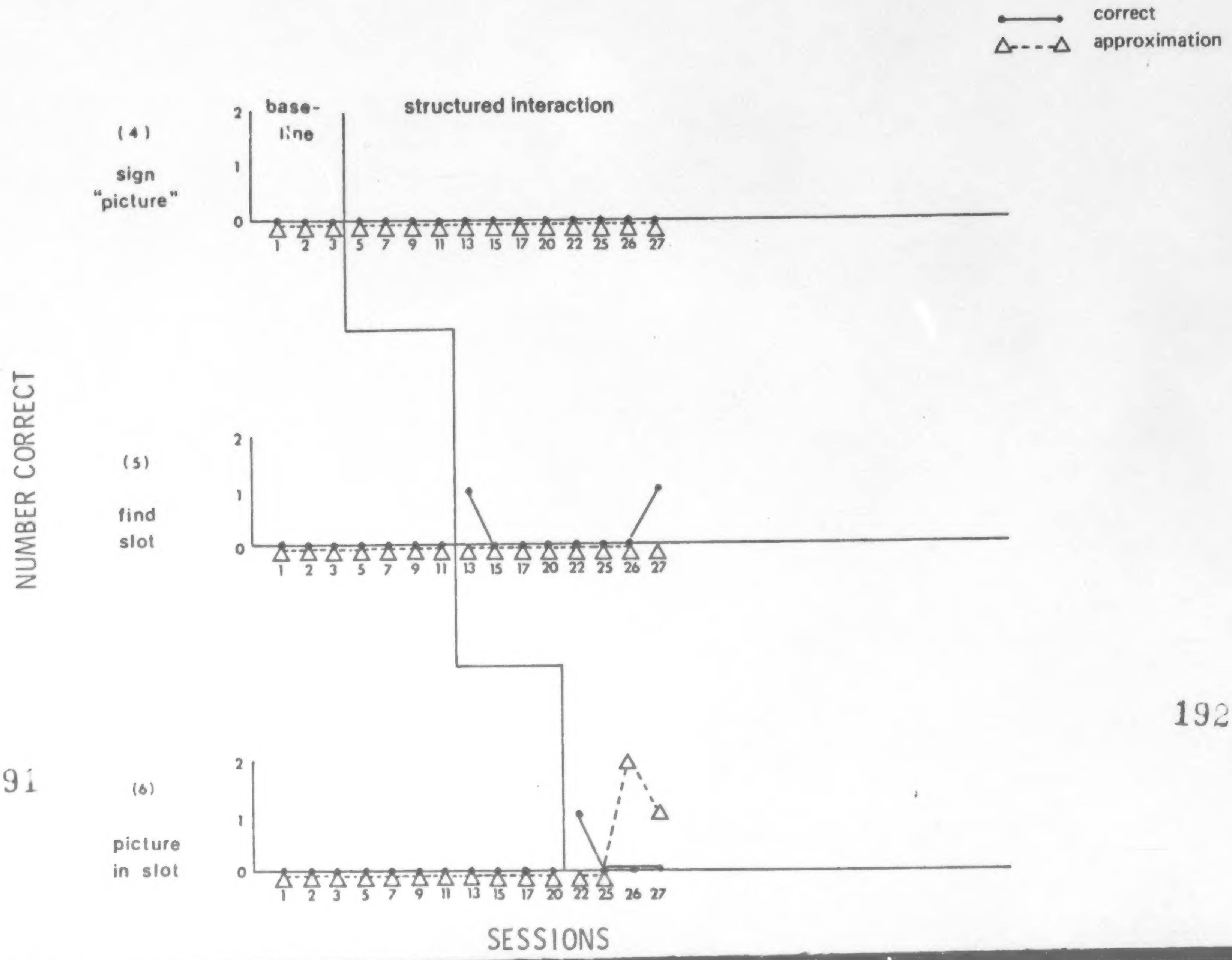


Figure 4. Bill's incidental learning of the three tasks that he observed Dave being directly instructed on. Incidental learning was assessed during the Probe component of the session.

Bill: INCIDENTAL LEARNING-INDIVIDUAL PROBE



The difference in the demonstration of incidental learning between the two students in the dyad is very significant; i.e., Bill showed almost no evidence of incidental learning, while Dave's performance was much more dramatic. Examination of the students' acquisition of their own instructional programs might shed some light on this. Dave demonstrated a rapid acquisition of all of his instructional programs compared to Bill's very slow and inconsistent learning of his own three tasks. (These data are available from the author). It seems to follow that if a student is having difficulty acquiring skills that are being directly taught to him, that incidental learning would be even more deficient.

Dave's performance of "approximations" on the Terminal Objective probes and "corrects" on Individual Step probes further confirms incidental learning. This implies that what Dave acquired was contingent on what he saw. In the Terminal Objective probe he did not have the opportunity to observe the terminal objective and therefore could only guess or approximate the skill. However, in the Individual Step probe, Dave was able to observe the cues and correction procedures of the behavior that he was to later demonstrate.

The findings of this study have several implications for application of group instruction to classrooms for severely handicapped students. In view of how much incidental learning can occur, the teacher should consider how each student's objectives relate to the other student's objectives in the dyad or group. Arrangements can be made to group students who have objectives in common. Grouping a student with an appropriate behavior with a student who demonstrated an incompatible behavior would be another possible arrangement. Structured interactions could then be implemented at those parts of the sequence where learning would be most advantageous; e.g., before or after the incompatible behavior.

Results of the present research suggest the need for further inquiries into the level of assistance needed to improve a skill learned incidentally to criterion level. Although it seems that Bill did not benefit from the opportunity to incidentally learn from his peer, it is possible that this exposure might decrease the amount or intensity of direct training needed to acquire those skills. This might be an especially relevant feature for nonimitative students, and worthy of future research efforts.

Reference Notes

1. Sullivan, K. M. *The effects of social interaction among handicapped and non-handicapped preschool children*. Unpublished master's thesis, University of Kansas, 1977.
2. Yelenik, G. *Effect of peer model training on the acquisition of an independent living skill by two severely and multiply handicapped students*. Unpublished thesis, University of Kansas, 1980.
3. Ruggles, T. R., & LeBlanc, J. M. *Variables which affect the effectiveness of group training procedures designed for children with learning problems*. Unpublished manuscript, University of Kansas, 1979.

References

Alberto, P., Jobes, N., Sizemore, A., & Doran, D. A comparison of individual group instruction across response tasks. Journal of the Association for Severely Handicapped, 1980, 5, 285-293.

Bandura, A. Aggression: A social learning analysis. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1973.

Barrett, B. H. Communication and the measured message of normal behavior. In R. L. York and E. Edgar (Eds.). Teaching the severely handicapped. Vol. IV. Seattle, WA: American Association for the Education of the Severely/Profoundly Handicapped, 1979, 301-318.

Brown, F., Holvoet, J., Guess, P., & Mulligan, M. The individualized curriculum sequencing model (III): Small group instruction. Journal of the Association for Severly Handicapped, 1980, 5(4), 352-367.

Brown, L., Nietupski, J., & Hamre-Nietupski, S. Criterion of ultimate functioning. In M. A. Thomas (Ed.) Hey, don't forget about me. Reston, VA: The Council for Exceptional Children, 1976, 2-15.

Doll, E. Vineland Social Maturity Scale. Circle Pines, Minn.: American Guidance Service, 1965.

Favell, J. E., Favell, J. E. & McGimsey, J. F. Relative effectiveness and efficiency of group versus individual training of severely retarded persons. American Journal of Mental Deficiency, 1978, 83, 104-109.

Guralnick, M. J. The value of integrating handicapped and nonhandicapped preschool children. American Journal of Orthopsychiatry, 42, 1976, 236-245.

Kohl, F., Wilcox, B., & Karlan, G. Effects of training conditions on the generalization of manual signs with moderately handicapped students. Education and Training of the Mentally Retarded, 1978, 13(3), 327-335.

Mithaug, D. E., & Wolfe, M. S. Employing task arrangements and verbal contingencies to promote verbalizations between retarded children. Journal of Applied Behavior Analysis, 1976, 9, 301-314.

Orelove, F. P. Acquisition of incidental learning in moderately and severely handicapped adults. Education and Training of the Mentally Retarded, 1982, 17(2), 131-136.

Sailor, W., & Mis, B. TARC Assessment System. Lawrence, KS.: H and H Enterprises, 1975.

Snyder, L., Appoloni, T., & Cooke, T. P. Integrated settings at the early childhood level: The role of non-retarded peers. Exceptional Children, 1977, 43, 262-266.

Storm, R. H. & Willis, J. H. Small-group training as an alternative to individual programs for profoundly retarded persons. American Journal of Mental Deficiency, 1978, 83, 283-288.

Stutsman, R. Merrill Palmer Scale of Mental Tests. C. H. Stoelting Co., 1931.

Group Instruction and Incidental Learning in Learners with Severe Mental Retardation

Jerry Rea
Kansas State School for the Deaf
Olathe, KS

Jennifer Holvoet, Ph.D.
University of Kansas
Lawrence, KS

Nancy Schussler, M.Ed.
University of Kansas
Lawrence, KS

1983

This paper was prepared as part of the work pursuant to the "Kansas Individualized Curriculum Sequencing Project" funded as a Handicapped Children's Model Project, U.S. Office of Education (G00-800-1721) from 9-1-80 to 9-1-83.

There have been two approaches to research regarding the function of group instruction with severely, multiply handicapped students. The first approach compared group instruction to 1:1 instruction in terms of efficient use of teacher time (c.f., Alberto, Jobes, Sizemore, & Doran, 1980; Favell, Favell, & McGimsey, 1978; Kohl, Wilcox, & Karlan, 1978; Storm & Willis, 1978). The other approach looked at whether group instruction provides unique opportunities for students to learn skills that are difficult to teach in a 1:1 instructional format (Brown & Holvoet, 1982; Orelove, 1982). These latter investigators have been particularly interested in the fact that group instruction provides a format in which incidental learning might take place. Incidental learning is defined as learning that takes place without direct instruction. If a group were composed of students who were being taught different skills, it is possible that the students might learn one another's skills through the mechanism of incidental learning. Bandura (1973), however, points out that incidental learning is not only a function of being given the opportunity to observe others, but is also a function of the degree to which a person attends to the important features of what is observed.

Since severely handicapped learners often show deficits in attending to others, it may be necessary to manipulate the environment or teaching situation in some manner to ensure that the students attend to one another's behavior during group instruction. One way of manipulating the teaching situation is simply to cue the student to attend to the relevant components of another student's performance (e.g., "Watch John put the mop in the bucket."). Several investigators (Guralnick, 1976; Sullivan, Note 1; Yelenik, Note 2) have been able to increase the efficacy of peer-modeling in small groups of severely handicapped students by using this method.

Another, and perhaps less intrusive technique, might be to structure the teaching situation so that it includes a systematic interaction component. This means that the teaching situation should be structured in such a way as to demand interaction between the participants. This has generally been accomplished by engineering situations where both participants must actively cooperate to reach a mutually reinforcing goal. For example, Ruggles and LeBlanc (Note 3) found that nonhandicapped children learned more from one another when there was sytematic interaction between the students. Mithaug and Wolfe (1976) found that structured peer interaction increased verbal behavior in handicapped students, though there was no actual incidental learning.

Brown, Holvoet, Guess, & Mulligan (1980) discussed using an intersequential model of group instruction with the severely handicapped population. This model involved having students act as antecedents and/or consequents to each other's behavior. The effect on incidental learning of this model was tested by Brown & Holvoet (1982). These authors found that one of two students showed incidental learning when he was required to provide edible consequences for the other sudent's correct responses. The other student learned neither the directly taught nor the incidental skills.

This study is a systematic replication of the Brown and Holvoet (1982) study. This study proposes to investigate whether interactions based on antecedent events (i.e., having the student give the cue or the materials needed for the other student to do the task) would be as effective in eliciting incidental learning as interactions based on consequent events (i.e., having the students give each other edibles for correct performance).

METHOD

Participants

Four students with severe mental retardation were selected from a secondary public school classroom for students with severe handicaps. The criteria for selection were: 1) the student had not participated in any research related to group instruction during the past two school years; 2) the student could imitate actions performed by an adult; and 3) the student was available to participate in research at the same time as the researcher.

The first participant, Don, was an 18-year old male who scored in the 70th percentile on the TARC Assessment Inventory for Severely Handicapped Children (Sailor & Mix, 1975). Don had no physical disabilities and lived at home with his natural parents. He sometimes exhibited non-purposive (stereotyped) vocalization, both in the classroom and during research sessions. Though he was characterized as often showing noncompliant behavior in the classroom, his compliance was excellent during the research sessions.

The second participant, Mary, was a 16 year-old female who scored in the 50th percentile on the TARC Assessment Inventory for Severely Handicapped Children. She exhibited no physical disabilities and resided at a state institution for the mentally retarded. She sometimes exhibited non-purposive (stereotyped) head movements in both the classroom and research sessions. She was characterized as a behavior problem by classroom staff due to periodic intense tantrums, but these were not a problem during the research sessions.

David, the third participant, was a 19 year-old male who lived at home with this natural parents. He socred in the 25th percentile on the TARC Assessment Inventory for Severely Handicapped Children. He was unable to walk due to severe malformations of the feet. He also showed a lack of strength and fine motor coordination in the upper extremities, particularly on the right side. He was, however, able to push his own wheelchair and to manipulate objects on command. He often exhibited stereotyped movements of the arms and head during the research sessions, and made stereotyped vocalizations regularly. On occasion, he hit his head with his elbow and/or bit himself, but these behaviors were rare enough, and of a mild enough intensity, that they did not interfere with instruction.

Ben, the fourth participant in the study, was a 19 year-old male who lived in a state institution for the mentally retarded. He scored in the 45th percentile on the TARC Assessment Inventory for Severely Handicapped Children. Ben was diagnosed as having athetoid cerebral palsy, with moderate involvement of all four extremities. He was confined to a wheelchair, but was able to grasp and manipulate objects with his hands if given sufficient time. He exhibited no stereotyped behavior and was very compliant both in the classroom and research sessions.

Construction of dyads. The participants were divided into dyads for the duration of the study. The first dyad was composed of Don and Mary, while the second dyad was composed of David and Ben. Don and Mary were placed in the same dyad because neither had a physical handicap that would require an alteration in the topographies of the tasks. David and Ben, because of their physical disabilities, both required altered response topographies in order for them to perform the tasks.

Setting

Training and probe sessions were conducted in an area which measured approximately 20' x 45'. This room was divided into two equal parts with one half used for occupational therapy training. The sessions were conducted in the occupational therapy area. Adaptive equipment was in this area, but no other staff or students were present. Once a week, music training was conducted in the other section of the room at the time the research sessions were in progress. This did not seem to be distracting to the research subjects. The subjects were seated in desks and chairs (or wheelchairs) adjacent to each other, while the experimenter was seated opposite them.

Tasks/Materials

Skills were chosen in terms of the needs of the students in both dyads. For instance, if the students in one dyad needed to learn to count, but that task was not appropriate for a student in the other dyad, it was not included in the study. But, if that task was appropriate for all the students, then it was directly taught to one of the students in each group.

In order to determine which skills were needed by all the students, a pretest was conducted with all four participants. This pretest was composed of ten different tasks. The students' teachers had chosen from a list of thirty tasks, ten that were appropriate for the student(s). The pretest was conducted in a 1:1 setting and the student was given two trials of each task. No training was implemented during the pretest, but the student was praised, on an intermittent basis, for "working," "sitting nicely," "paying attention," and "trying." The pretest was implemented on two consecutive days to be sure that performance stayed at the same level. From these ten tasks, the experimenters chose four tasks that all four of the students failed. Failure was defined as the student making no correct, or approximately correct, response on any of the trials of that task. The following is the list of instructional programs that were directly taught to the specific students:

Don

1. Time Telling
2. 1:1 Correspondence

Mary

1. Match to Sample
2. Letter Box Assembly

David

1. Match to Sample
2. Letter Box Assembly

Ben

1. Time Telling
2. 1:1 Correspondence

Daily data were recorded on each student's acquisition of these instructional programs. A teaching strategy of modeling and physical guidance was used to correct student errors during the instructional sessions, and correct responses resulted in praise.

Measurement of Incidental Learning

Each student in the dyad was given an opportunity at the end of the instructional session to demonstrate whether s/he had learned the other student's instructional tasks (probe trials). There was no direct instruction during the probe trials, but reinforcement was provided for a correct response during these probe trials in order not to "extinguish" correct responding. The probe trials were administered twice for each instructional task and probes were alternated between the students in the dyad.

During a probe trial, the materials were presented and a general cue was given. The student's response was recorded as + (correct), - (incorrect or no response), or A (approximation of a correct response).

Procedures

Incidental learning was measured in a multiple baseline (Herson & Barlow, 1976) format across untrained tasks for each student in the dyad.

Baseline Condition. The baseline condition consisted of no systematic interactions between the members of the dyad. The experimenter alternated instruction between the two students on their own instructional tasks without requiring any interaction between the students. At the end of the instructional session, the experimenter conducted a probe session. The experimenter allowed each student two attempts to do the other student's tasks. This was done to see, for example, if Don was observing and learning Mary's tasks and if Mary was learning Don's tasks.

Systematic Intersequential Interaction Condition. During the Systematic Interaction condition the members of the dyad were required to give each other the materials needed to do the tasks and to remove the completed products during the instructional session. For example, Mary was cued (if necessary) to give Don the Time-telling materials. When this was done, the experimenter gave Don the cues and prompts needed to ensure Don's correct performance. Then Mary was cued to remove the time-telling materials. At the end of the instructional session the experimenter conducted a probe session (e.g., gave Mary the opportunity to do Don's Time-telling task). The experimenter gave each student two opportunities to do the other student's tasks. These probe sessions were conducted exactly like those in the Baseline condition.

To control for order effects, the tasks included in both the instructional and probe sessions were alternated between students within each of their respective dyads.

Reliability

Interobserver reliability data were obtained during the Baseline and Systematic Intersequential Interaction conditions. These data were taken during the probe sessions. The observer sat approximately three feet from the students and the experimenter and independently recorded whether the student's responses were correct, incorrect, or approximations. Reliability measures for each task were calculated by dividing the number of agreements by the number of agreements plus disagreements and multiplying by 100.

RESULTS

Reliability

Reliability data were obtained on 27 occasions during the Probe sessions where incidental learning was measured. Twenty-six of these observations resulted in 100% interobserver agreement; 75% reliability was obtained during the other session.

Incidental Learning

Figures 1-4 illustrate, in graphic form, the incidental learning of the students in the two dyads. Figures 1 and 2 illustrate the performance of Don and Mary (Dyad 1) respectively, while Figures 3 and 4 illustrate Ben's and David's (Dyad 2) performance. Four graphs are included in each figure. The graphs on the left side of the figure represent two different tasks and show the number of trials in which the student made a correct response. The graphs on the right side of the figure show the number of trials in which the student made an approximation of the correct response. Thus, it can easily be seen whether requiring a systematic interaction resulted in complete incidental learning (correct trials) and/or partial incidental learning (approximation trials). It should be pointed out that though the pretest data are not graphed, all students showed a zero level of performance on these tasks when tested individually. Therefore, any performance above the zero level during the Baseline condition (where group training was conducted without systematic interaction), can be interpreted as evidence that incidental learning had occurred.

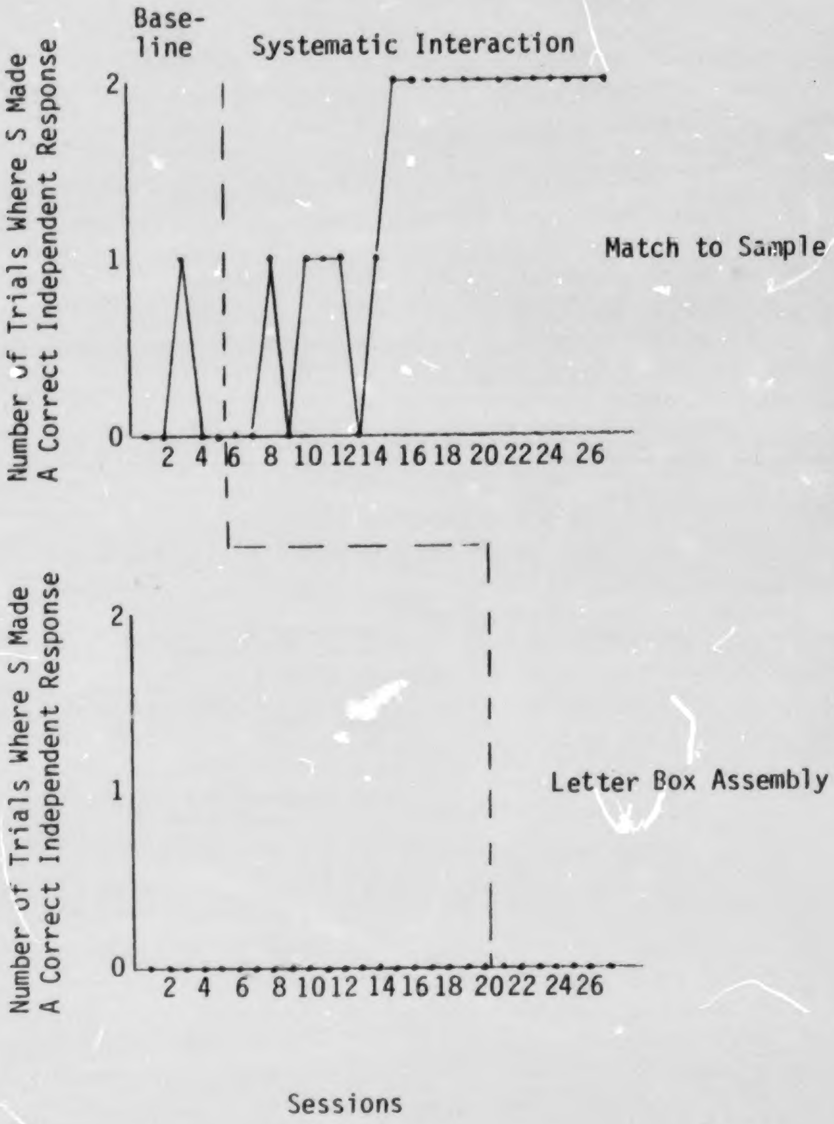
Insert Figures 1-4 about here

Figure 1 depicts Don's performance. It can be seen that there was some incidental learning (both at the correct and approximated level) during the Baseline condition on the Match-to-sample task. Incidental learning continued to be manifested on this task during the Systematic interaction condition, and complete incidental learning was demonstrated on the tenth Systematic Interaction session. No approximations were seen in the Systematic interaction condition, as would be expected with a high level of correct responding.

On the Letter Box Assembly task, Don showed only approximated incidental learning in both conditions. No difference in performance was seen across conditions with this task.

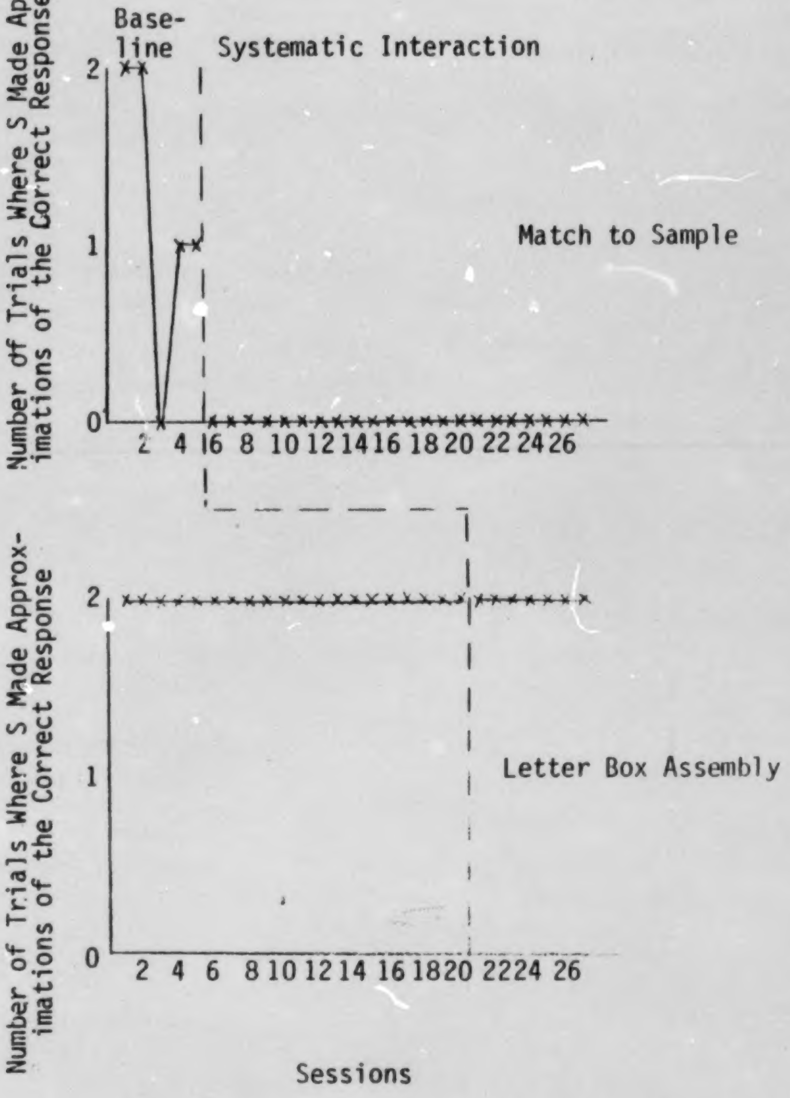
Figure 2 depicts Mary's performance. With the exception of one session, she shows no incidental learning of the time-telling task in either condition. On the 1:1 Correspondence task she showed immediate

Figure 1. Don's incidental learning on two tasks. The left-hand graphs represent the number of trials on which Don made a correct response. The right-hand graphs represent the number of trials on which Don made an approximation of the correct response.



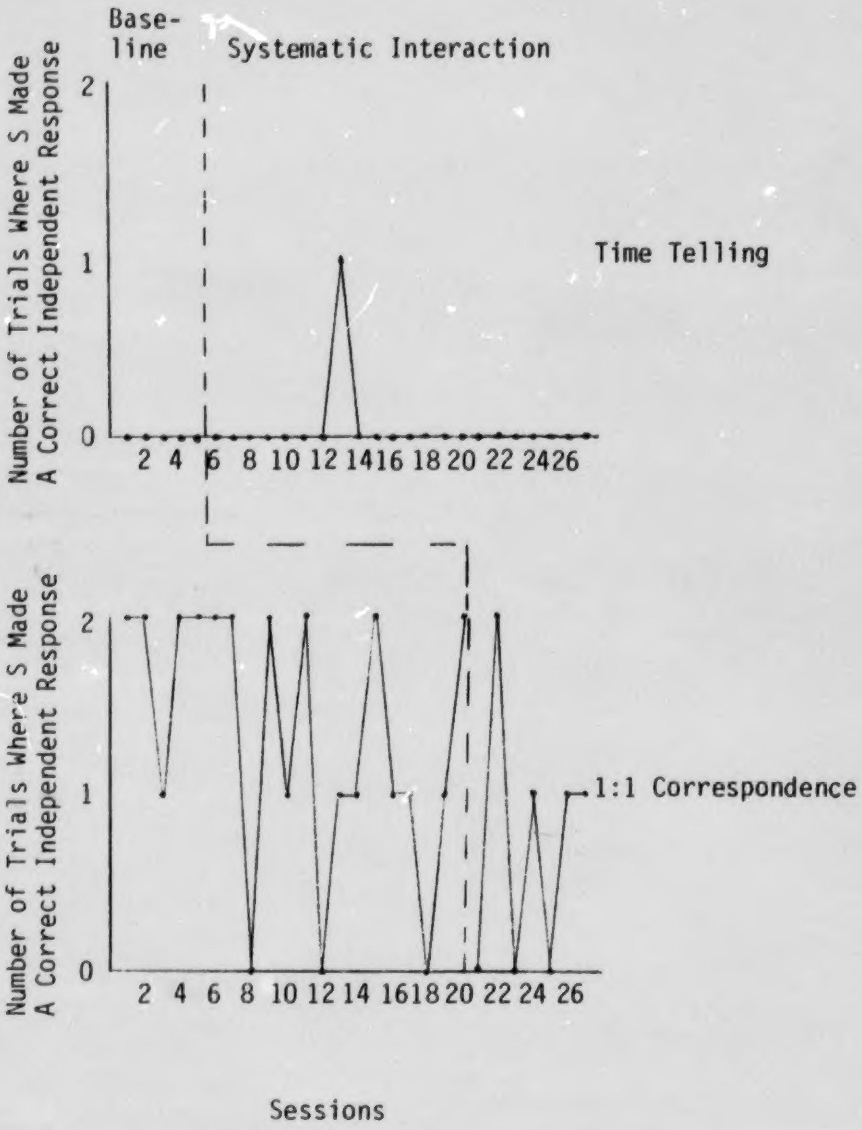
Base-
line
Systematic Interaction
Number of Trials Where S Made
A Correct Independent Response
2
1
0
2 4 6 8 10 12 14 16 18 20 22 24 26
Match to Sample
Number of Trials Where S Made
A Correct Independent Response
2
1
0
2 4 6 8 10 12 14 16 18 20 22 24 26
Letter Box Assembly
Sessions

Don



Base-
line
Systematic Interaction
Number of Trials Where S Made Approx-
imations of the Correct Response
2
1
0
2 4 6 8 10 12 14 16 18 20 22 24 26
Match to Sample
Number of Trials Where S Made Approx-
imations of the Correct Response
2
1
0
2 4 6 8 10 12 14 16 18 20 22 24 26
Letter Box Assembly
Sessions

Figure 2. Mary's incidental learning on two tasks. The left-hand graphs represent the number of trials on which Mary made a correct response. The right-hand graphs represent the number of trials on which Mary made an approximation of the correct response.


Base-
line
Systematic Interaction
Number of Trials Where S Made A Correct Independent Response
2
1
0
2 4 6 8 10 12 14 16 18 20 22 24 26
Time Telling
Number of Trials Where S Made A Correct Independent Response
2
1
0
2 4 6 8 10 12 14 16 18 20 22 24 26
1:1 Correspondence
Sessions

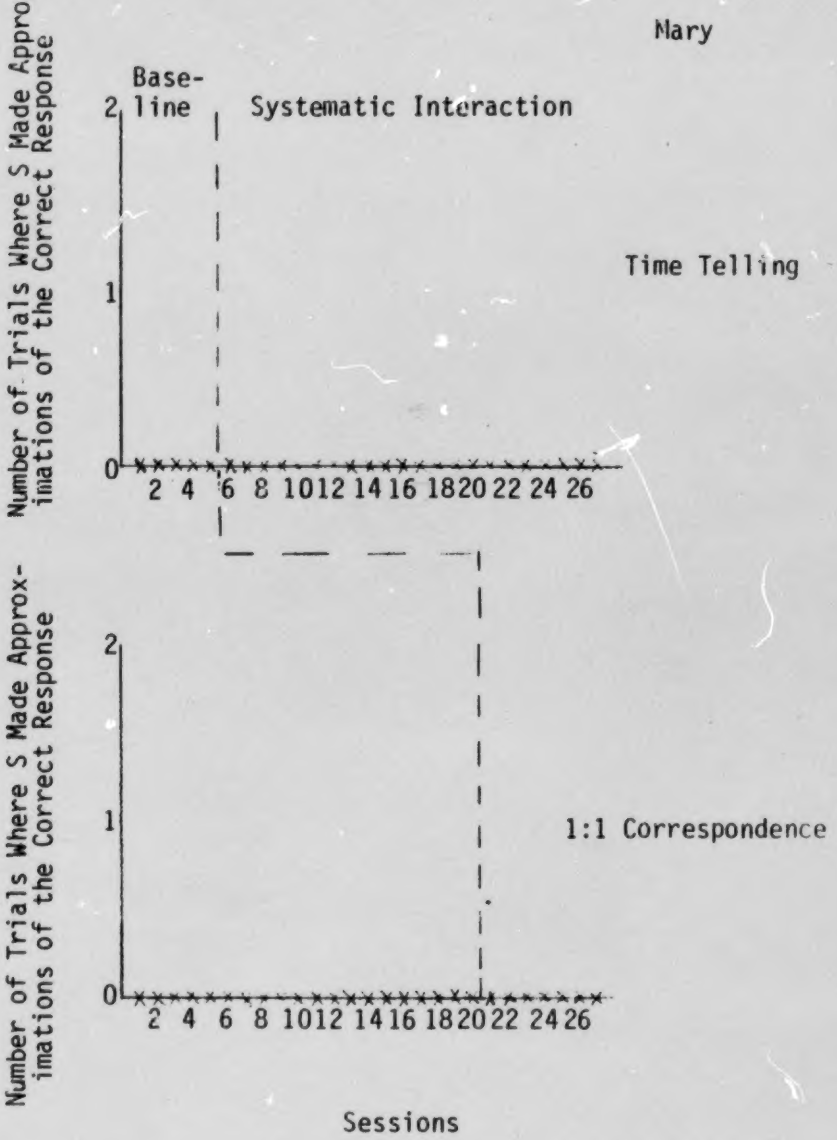
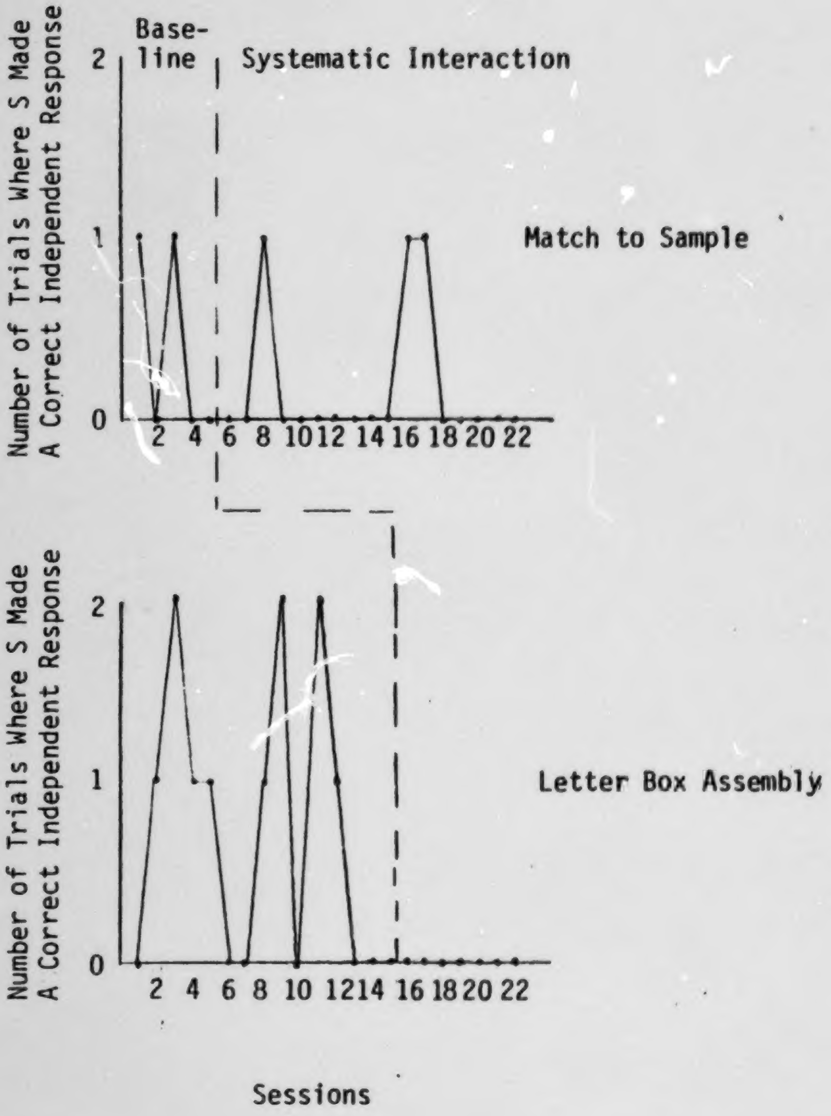

Mary
Base-
line
Systematic Interaction
Number of Trials Where S Made Approximations of the Correct Response
2
1
0
2 4 6 8 10 12 14 16 18 20 22 24 26
Time Telling
Number of Trials Where S Made Approximations of the Correct Response
2
1
0
2 4 6 8 10 12 14 16 18 20 22 24 26
1:1 Correspondence
Sessions

Figure 3. Ben's incidental learning on two tasks. The left-hand graphs represent the number of trials on which Ben made a correct response. The right-hand graphs represent the number of trials on which Ben made an approximation of the correct response.


Base-
line
Systematic Interaction
Number of Trials Where S Made A Correct Independent Response
2
1
0
Match to Sample
2 4 6 8 10 12 14 16 18 20 22
Number of Trials Where S Made A Correct Independent Response
2
1
0
Letter Box Assembly
2 4 6 8 10 12 14 16 18 20 22
Sessions

Ben

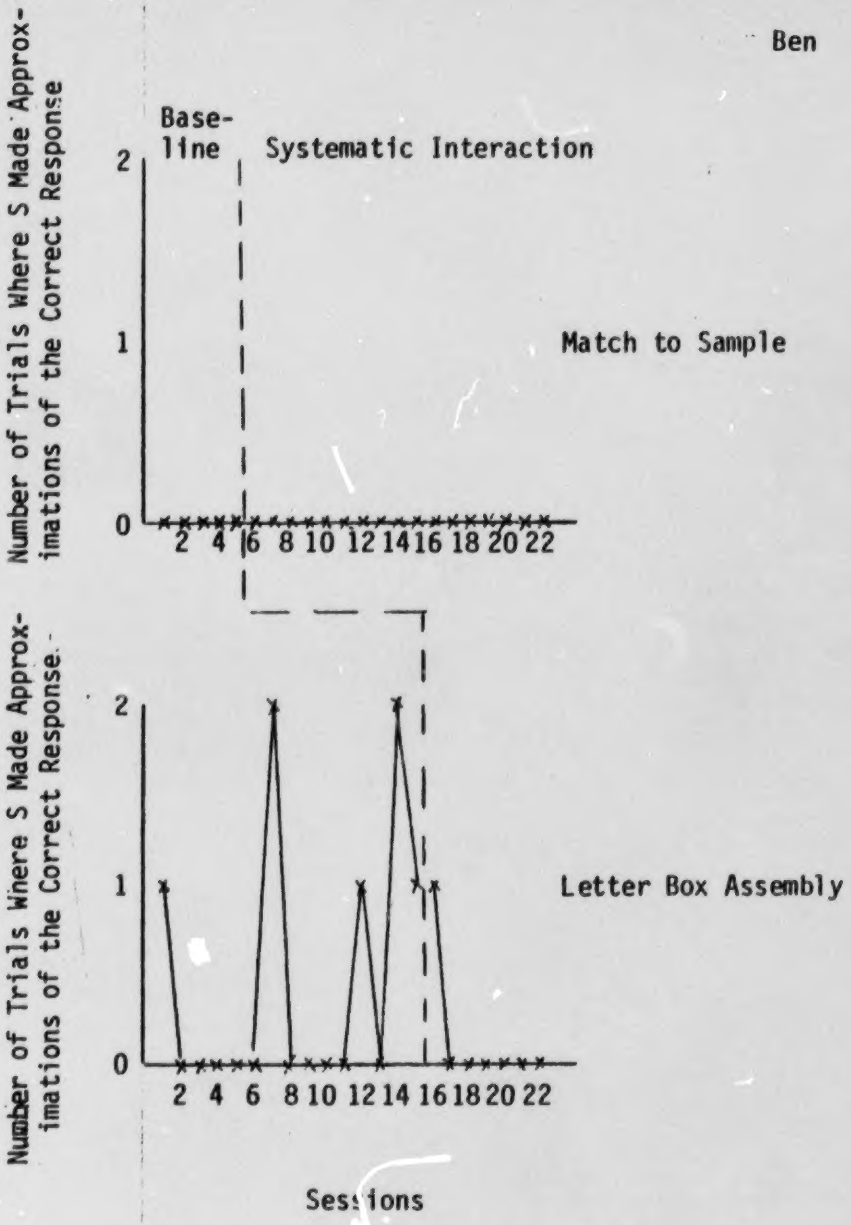
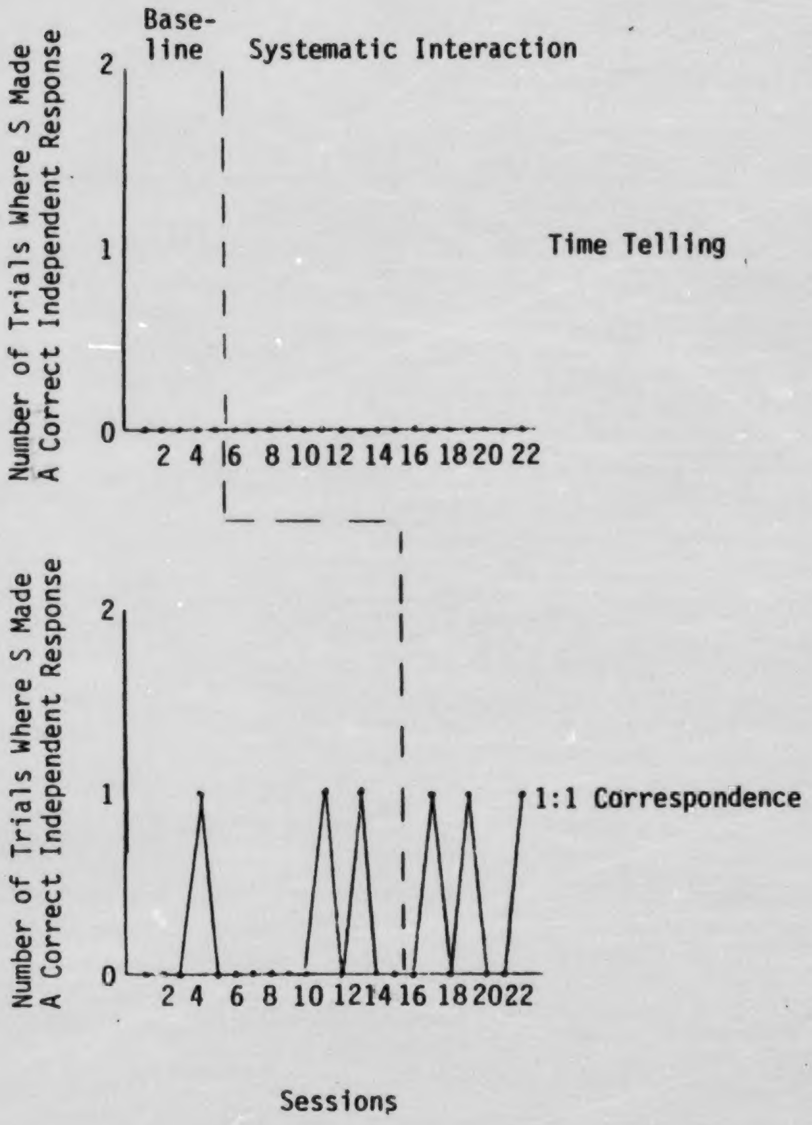
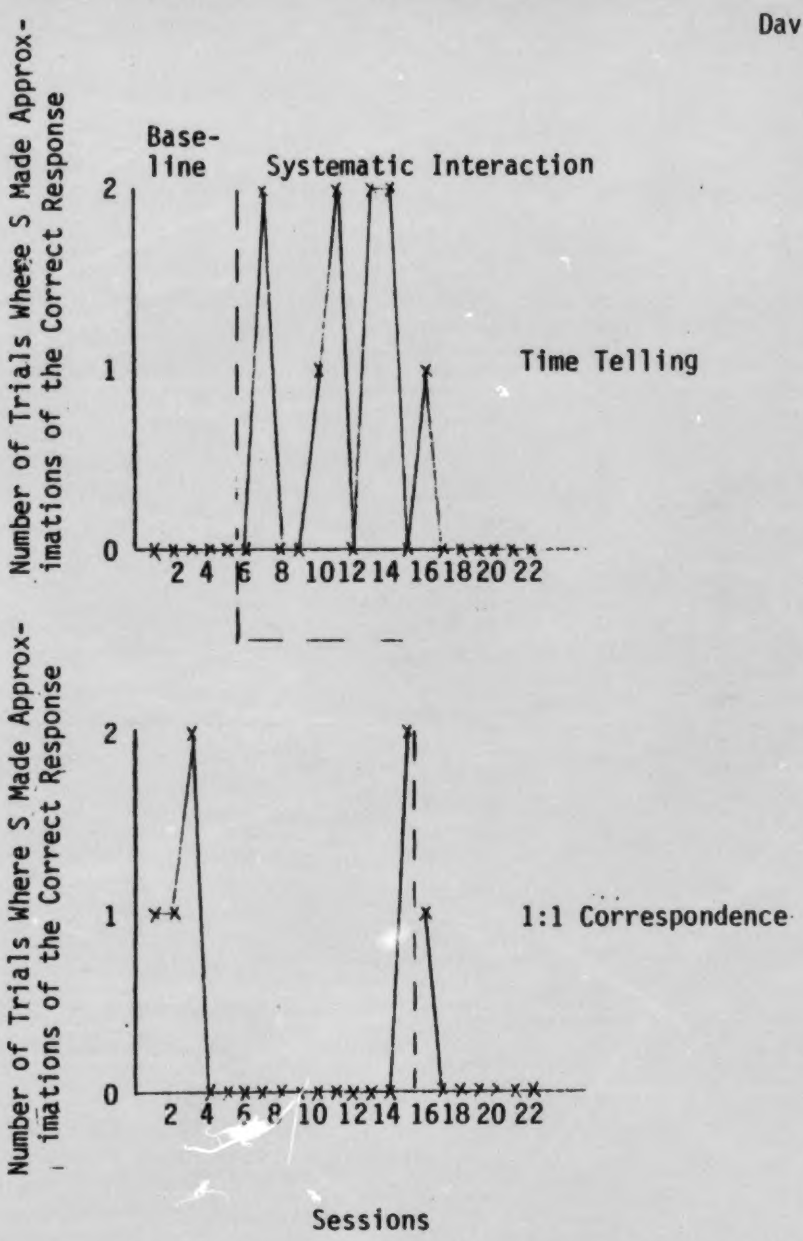

Base-
line
Systematic Interaction
Number of Trials Where S Made Approximations of the Correct Response
2
1
0
Match to Sample
2 4 6 8 10 12 14 16 18 20 22
Number of Trials Where S Made Approximations of the Correct Response
2
1
0
Letter Box Assembly
2 4 6 8 10 12 14 16 18 20 22
Sessions

Figure 4. David's incidental learning on two tasks. The left-hand graphs represent the number of trials on which Don made a correct response. The right-hand graphs represent the number of trials on which Don made an approximation of the correct response.

David


Base-
line
Systematic Interaction
Number of Trials Where S Made A Correct Independent Response
2
1
0
2 4 6 8 10 12 14 16 18 20 22
Time Telling
1:1 Correspondence
Sessions


Base-
line
Systematic Interaction
Number of Trials Where S Made Approximations of the Correct Response
2
1
0
2 4 6 8 10 12 14 16 18 20 22
Time Telling
1:1 Correspondence
Sessions

and correct incidental learning during the first seven sessions of the Baseline condition. Her performance after that point was highly variable across sessions. This same variable pattern was seen when Systematic interaction was required, but the overall performance level appeared to be somewhat lower in this condition. Mary never showed any approximated responses in either condition of either task.

Figure 3 depicts Ben's incidental learning. He made some correct responses during both conditions of the Match-to-sample task. His performance, however, appeared to be lower when Systematic interaction was introduced. He made no approximated responses on this task in either condition.

On the Letter Box Assembly task, Ben showed a high degree of both correct and approximated incidental learning during the Baseline condition. With only two exceptions in this condition, he made a correct or approximated response on at least one of the two trials. His performance, however, showed a marked decrease when the Systematic interaction condition was in progress.

Figure 4 depicts David's performance. On the Time-telling task, David showed no incidental learning during the Baseline condition. He did, however, show some approximated responses on several occasions during the Systematic interaction condition.

On the 1:1 Correspondence task, David showed some incidental learning that varied in a fairly regular pattern across sessions. This pattern of showing some learning in one session and none on the next was seen in both conditions.

DISCUSSION

These results were exciting though somewhat different from what was anticipated. The data indicate clearly that every student showed some incidental learning. Although only one of the students (Don) learned a task completely (i.e., several consecutive sessions with both trials correct), all the students made approximated and/or correct responses across several sessions. Don, David, and Ben showed incidental learning on both their tasks; Mary showed it on only one. Interestingly enough, on most of the tasks where incidental learning was demonstrated, this learning was manifested in the Baseline condition. Since these students had shown zero levels of performance on these tasks in 1:1 situations (Pretest), it appears that these students were able to incidentally learn simply by being placed in a group situation. This finding was not anticipated because most severely handicapped learners (these subjects included) have needed very powerful techniques to achieve success in situations where direct instruction was offered. Therefore, most professionals and teachers have made the assumption that these students would not be likely to learn tasks that were not directly taught (i.e., would have problems in incidental learning). This assumption was further bolstered by the fact that many individuals with severe handicaps have not learned the skills that are routinely demonstrated by others in their environment.

The fact that these students were able to learn incidentally may be related to four factors. First of all, these students all had the skill of motor imitation and routinely were able to learn when the teacher used a Demonstration prompt. It would appear logical that students who are routinely imitative would be more likely to learn incidentally than those who are not imitative.

Second, it is possible that the situation encouraged incidental learning. These students had a long history of being taught while seated at a desk with an instructor presenting materials and cues. Thus, they might be "primed" to attend even when the materials and cues were presented to another student.

A third possibility is that most severely handicapped learners are capable of incidental learning, but the current teaching techniques are such that the learners either are not required to demonstrate incidental learning or do not have a chance to incidentally learn. The authors have observed that the bulk of teaching with this population still seems to utilize 1:1 direct instruction or groups where students are taught the same skills. Even in those rare instances where group members are taught different skills, incidental learning is never assessed.

A fourth possibility is that most severely handicapped learners are capable of, and engage in, incidental learning but it escapes the awareness of professionals because the severely handicapped learner may only approximate the correct response and/or may make the correct response only occasionally.

It would appear that more effort needs to be made, both in research settings and in classrooms, to look at the phenomenon of incidental learning. If most students show some incidental learning, allowing them to work in diverse groups and demonstrate what has been learned could save the teacher quite a bit of instructional time. It is certainly easier to shape an approximated response or to increase the frequency of correct responses than it is to build a skill "from scratch."

The failure of the systematic intersequential interaction to enhance incidental learning in these student was, however, disappointing and was not in accord with the findings of Brown and Holvoet (1982). Only Don and David showed any improvement during the Systematic Interaction condition, and even they did not show the effect on all their programs. The failure of this intervention may be related to any or all of several factors. First, it may be that these students had already shown all the incidental learning they were capable of, and adding interaction could not improve it. Brown & Holvoet's (1982) student had showed no previous incidental learning while in the Baseline group condition, possibly because their student was less imitative. Thus, it might be that systematic interaction may prompt incidental learning in those students who do not spontaneously manifest it in a group situation; but it may not improve incidental learning that has already taken place.

Second, it is possible that handling the materials is simply not as powerful an interaction as providing the edible consequences for correct performance. The materials must be provided and removed regardless of the correctness of the response, but the edible consequences used by Brown & Holvoet (1982) were provided only at the point where the response was correct.

Third, the students observed their partner making lots of errors during the training session. It was noted by the experimenter and the observers that the students usually watched each other for only a few seconds after the presentation of materials. Thus, if the student being trained made an error in the initial response, this was what the other student watched. Only rarely did the students observe one another during the correction procedure. This meant that the students were more likely to observe error responses than correct responses, which in turn increased the likelihood that what would be incidentally learned was the error response. This, in fact, is exactly what happened in some instances;

an approximated incidental response was exactly the same error behavior that had been observed earlier with the target student. The use of a shorter correction procedure, an errorless procedure, or a dyad partner who knew how to do the task correctly all might have improved incidental learning. The presentation and removal of materials by the students instead of the teacher did not seem to increase the duration of observation and, thus, did not have an impact. Apparently these students must observe the process of completing a task; simply seeing the product in its uncompleted state and then seeing it in its completed state is not enough.

It appears that some severely handicapped students have the ability to incidentally learn if given the opportunity. This ability may depend on how well the student has acquired the skill of motor imitation and/or the structure of the environment. The enhancement of incidental learning in these students may depend on providing environemntal cues that ensure observation of the correct response.

Reference Notes

1. Sullivan, K. M. The effects of social interaction among handicapped and nonhandicapped preschool children. Unpublished master's thesis, University of Kansas, 1977.

2. Yelenik, G. Effect of peer model training on the acquisition of an independent living skill by two severely and multiply handicapped students. Unpublished Master's Thesis, University of Kansas, 1980.

3. Ruggles, T. R., & LeBlanc, J. M. Variables which affect the effectiveness of group training procedures designed for children with learning problems. Unpublished manuscript, University of Kansas, 1979.

References

Alberto, P., Jobes, N., Sizemore, A., & Doran, D. A comparison of individual group instruction across response tasks. Journal of the Association for the Severely Handicapped, 1980, 5, 285-293.

Bandura, A. Aggression: A social learning analysis. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1973.

Brown, F. & Holvoet, J. Effect of systematic peer interaction on the incidental learning of two severely handicapped learners. Journal of the Association for the Severely Handicapped, 1982, 7, 19-28.

Brown, F., Holvoet, J., Guess, P., & Mulligan, M. The individualized curriculum sequencing model (III): Small group instruction. Journal of the Association for the Severely Handicapped, 1980, 5, 352-367.

Favell, J. E., Favell, J. E., & McGimsey, J. F. Relative effectiveness and efficiency of group versus individual training of severely retarded persons. American Journal of Mental Deficiency, 1978, 83, 104-109.

Guralnick, M. J. The value of integrating handicapped and nonhandicapped preschool children. American Journal of Orthopsychiatry, 1976, 42, 236-245.

Hersen, M. & Barlow, D. H. Single-case experimental designs: Strategies for studying behavior change. New York: Pergamon Press, 1976.

Kohl, F., Wilcox, B., & Karlan, G. Effects of training conditions on the generalization of manual signs with moderately handicapped students. Education and Training of the Mentally Retarded, 1978, 13, 327-335.

Mithaug, D. E., & Wolfe, M. G. Employing task arrangements and verbal contingencies to promote verbalizations between retarded children. Journal of Applied Behavior Analysis, 1976, 9, 301-314.

Orelove, F. P. Acquisition of incidental learning in moderately and severely handicapped adults. Education and Training of the Mentally Retarded, 1982, 17, 131-136.

Sailor, W. & Mix, B. TARC assessment system. Lawrence, KS: H & H Enterprises, 1975.

Storm, R. H. & Willis, J. H. Small-group training as an alternative to individual programs for profoundly retarded persons. American Journal of Mental Deficiency, 1978, 83, 283-288.

Effects of an Assembly-Line Format on the Work Productivity of Retarded Students

Jennifer F. Holvoet, Ph.D.

Blane Brown, B. A.

Nancy Schussler, M.Ed.

Doug Guess, Ed.D.

University of Kansas
Lawrence, Kansas

1983

This paper was prepared as part of the work pursuant to the "Kansas Individualized Curriculum Sequencing Project" funded as a Handicapped Children's Model Project, U.S. Office of Education (G00-800-1721) from 9-1-80 to 9-1-83.

Vocational and prevocational training is an area of great concern to educators trying to meet the needs of severely handicapped adolescents. It has long been recognized that there are two separate issues that need to be addressed when preparing students for the workplace (Wehman & McLaughlin, 1980). The first issue is one of acquisition. Students need to acquire the skills necessary to hold a job, either in the community or in a sheltered workshop. This process involves identifying relevant skills and applying instructional techniques to teach the skills. The second issue is one of production. The student, having acquired the skill, then needs to use it consistently, and accurately at a rate considered acceptable by the workplace. Much effort has gone into addressing the first issue, with techniques being delineated that increase the likelihood that students with severe handicaps will learn marketable skills (Bellamy, Peterson & Close, 1975; Gold, 1972, 1976; Gold & Barclay, 1973; Irvin, 1976; Levy, 1975; Spooner & Hendrickson, 1976; Wehman & Hill, 1982). As these techniques to enhance acquisition were applied, it became obvious to some educators, researchers and theoreticians that the techniques were not sufficient to stimulate acquisition in all students. Thus, some researchers began to focus on variables outside the teacher-student interaction, such as task structure and the environment. From this work, the concepts of job enlargement (i.e., the opportunity to work on more complex tasks) and community-based training were discovered to be variables that could enhance acquisition (Bellamy, Sheehan, Horner & Boles, 1980; Clarke, Greenwood, Abramovitz & Bellamy, 1980; O'Neill & Bellamy, 1978; Revell, Arnold, Taylor & Zaitz-Blotner, 1982).

Techniques were also developed to address the issue of production rate. Most of these techniques focused on the addition of prompts (Bellamy, et al., 1975; Brown & Pearce, 1970; Kliebhahn, 1967; Renzaglia, Wehman, Schutz & Karan, 1976) and manipulation of consequences (Brown, Van Deventer, Perlmutter, Jones & Sontag, 1972; Rusch, Connis & Sowers, 1979; Schroeder, 1972). It has long been recognized that these techniques are usually successful in increasing production, but that they are difficult to "fade out" with many students. It has also been noted that many supervisors, whether in a sheltered workshop or a community-based placement, do not have the time to deliver prompts or reinforcers to a client unless they have been faded to a very thin schedule. Unfortunately, recognition of this problem has not, for the most part, stimulated researchers and theoreticians to look "outside" the student for an answer, as was the case with acquisition.

This study looks at one component of the workplace, the grouping of workers, and its effect upon the production rate of two severely handicapped students. These students were chosen to exemplify the two types of students identified by Wehman and McLaughlin (1980) as having chronic production problems: 1) the worker who has very slow motor behavior and often appears to be unresponsive to the commonly used workshop incentives; and 2) the worker who exhibits high rates of disruptive behavior. The groupings chosen for study were: a) individual work, and b) work done in a group where a more skilled student worked with the target student in an assembly-line format. It was hypothesized that a student with slow motor behavior who needed a lot of prompting from the supervisor (e.g., "Get busy") would maintain a higher production rate in the assembly-line condition than in the condition where the student worked alone because the assembly-line format contains several inherent prompts to continue working (e.g., being given an item to complete). A similar hypothesis was held for the student who had a high rate of disruptive behavior. It was thought that if the disruptive behavior was an attention-getting device the presence of a peer might satisfy the student's need to interact socially.

METHOD

Subjects

Four male students were selected from two public school classrooms serving adolescents with special needs. Two of the students (the target students) were selected from a classroom for severely handicapped individuals and the other two (helper students) were chosen from a classroom for moderately retarded individuals. These four students were divided into two groups; each consisting of one helper and one target student.

The two target students, Tim and Woody, were 22 and 17 years of age, respectively, and were classified as severely retarded using the WISC-R and adaptive behavior measures from tests such as the TARC Assessment Guide (Sailor & Mix, 1975). These students were described by their teachers as "problem workers" needing almost constant 1:1 attention in the vocational setting. Tim often attempted to leave the work setting, destroyed materials and refused to continue working unless closely supervised. Woody never refused to work, or attempted to leave; instead, he moved very slowly while working and often stopped working and engaged in stereotyped behavior unless constantly prompted and praised. Both students could follow the simple directions involved in the vocational tasks chosen and could do the tasks correctly in a 1:1 format. Generally, the only prompts needed during a session were prompts to "get busy" and to "hurry."

The students who served as helpers, Steve and Fred, were 17 and 18 years of age, respectively. They were classified as moderately retarded using the WISC-R and tests of adaptive behavior such as the TMR Performance Profile (citation). Both Steve and Fred were very social, had good verbal skills, had excellent motor skills, and were characterized by their teachers as good workers in both academic and vocational settings.

The students were paired on the basis of the times they were available to work. Tim and Steve formed one pair and Woody and Fred formed the other group.

Setting

The study took place in a large, well lighted classroom of a self-contained public school serving special students. The two pairs of students were studied at separate times. The students were seated across from each other at a 4' x 6' table while observers watched from approximately five feet away. Five other students and two teachers were seated at a nearby table working on academic and vocational tasks unrelated to the research.

Tasks/Materials

Three tasks were used in this study: Placing barrettes on display cards, packaging pegs in plastic bags, and assembling hospital kits. The students were taught to do different components of each task. In the Barrette task, the helper student was taught to fasten a bar-type barrette to an empty cardboard display card, and the target student was taught to fasten one barrette to a partially completed display card (i.e., one that had one barrette already on it). In the Packaging pegs task, the helper student was taught to separate eight pegs from a large pile of pegs and to put these eight pegs in a styrofoam cup. The target student was taught to take a styrofoam cup with pegs in it and empty it into a plastic sandwich bag. Large plastic pegs were used in this task. Assembly of the hospital kit required the helper student to place a glass in a small plastic bag, put the wrapped glass in a pitcher, and put a lid on the pitcher. The target student was required to put the pitcher/glass assembly and a soap dish in a rectangular basin. These materials were all made of heavy plastic. Exactly the same behaviors were required of the students regardless of the experimental condition.

Measurement

There were three dependent variables in this study and each was measured separately for each of the three tasks. The variable of most interest was "on-task" behavior. This was defined as sitting quietly at the work table while moving or manipulating the work materials according to each task's directions; receiving a prompt or reinforcer from the trainer, or if the student had completed the task, waiting in his chair quietly until given additional work. Destroying the materials, taking apart previously completed work, or shaking the materials back and forth in a stereotypic manner were examples of "off-task" behavior. Each task was observed for 10 minutes. A time-sampling system, where the student was observed at the end of every 30 second interval, was used to record on-task behavior. The experimenter marked a + if the student was on-task at the time of the observation, and a - if the student was off-task. At the end of the session, the experimenter calculated the percent of intervals the student was on-task for each task and graphed the results.

The second variable, number of items completed, was measured at the end of each 10 minute session. The experimenter counted the number of items the student had completed during the session. Items were counted as being completed if they had been put in the bin for completed projects, even if they were incorrectly done. For example, a student was counted as having completed a package of pegs, if it had been put in the "out bin", even if the package only had 5 pegs in it instead of 8.

The third variable, number of items completed correctly, also was measured at the end of the 10 minute session. Items were counted as being correctly completed when the item was completed as specified in the task analysis and had been placed in the "out bin." For example, a package of pegs was only counted as being correct if it has exactly 8 pegs in it and was shut completely.

Experimental Design

A withdrawal of treatment (ABA) contained within a multiple baseline across tasks design (Herson & Barlow, 1976) was used to demonstrate experimental control with each student.

The students' work behavior was compared under two conditions. In the baseline condition, the helper and target students worked independently. At no time during this condition, did the two students work on the same task during the same 10 minute interval. For example, during the first 10 minutes of the session Tim might work on the Barrette task, while Steve worked on the Hospital Kit assembly. Then, during the second 10 minutes, Tim might work on Packaging pegs, while Steve worked on the Barrette task. In the last 10 minutes, Tim might work on the Hospital Kit assembly, while Steve worked on Packaging pegs. In this condition, designated the ALONE condition, interaction between the two students was not expected or encouraged.

In the second condition, the helper and target student worked together on the same task during each 10 minute interval. The helper student always began the task (e.g., put one barrette on the display card), then handed the item to the target student for completion. Thus, some interaction between the students was programmed in this condition. This condition was designated the ASSEMBLY-LINE condition.

Procedures

Alone condition. In the Alone condition, the helper was seated opposite the target student at the work table. The experimenter gave the helper student the materials for one task, and the target student the materials for a different task and showed each how to do their task correctly. After the demonstration had been given to both students, the experimenter stepped away from the table and told the students to begin working. The experimenter then went to an adjacent table and began collecting the time-sample data. The experimenter did not interact with the students until the end of the first five minutes. At this point, the experimenter praised the students individually if they were working or prompted them to "get busy" if they were not. If a student had completed all the materials given to him, the experimenter gave him more materials at this time. The experimenter continued to take "on-task" data during the time he was engaged in praise, prompting or setting up materials. After this scheduled interaction occurred, the experimenter returned to the adjacent table and continued collecting data. When 10 minutes had elapsed from the beginning of the session the experimenter stopped taking data, went to the work table, removed the unused materials, counted the number of items that had been completed and checked each item to see if it had been completed correctly. The students were then given feedback on how many items they had done correctly that day. The experimenter then presented each student with a different set of materials and proceeded as above. This procedure was repeated a third time after the students had completed their second tasks.

During the Alone condition, the students did not pass any materials to one another. If a student completed a task before the structured interaction with the experimenter was scheduled, he was to sit quietly (and was counted as being on task) until the experimenter gave him more work to do or the session ended. If the students talked to one another during the Alone condition, the experimenter ignored it and scored them as off-task only if they also had stopped working.

Assembly-line condition. In the Assembly-line condition, the students were seated across from one another at the work table. The materials for a task were placed in front of the helper student and he was shown how to do his portion of the task. He was instructed that when he had completed his part of the work, he was to pass the partially completed item to the other student, then begin working on another item. The experimenter then showed the target student how to do his part of the work. When the demonstrations were completed, the teacher moved away from the work table and told the students to begin working. The rest of the session followed the procedure outlined in the Alone condition.

It should be pointed out that the work expected of the students was exactly the same in both conditions. For example, in the Barrette task, the objective for the helper student was to put one barrette on the display card, whereas the objective for the target student was to put one barrette on a partially completed display card. Thus, in the Alone condition, the target student was given partially completed display cards to work on (i.e., the display cards had one barrette on them), but the helper student was given empty display cards. In the Assembly-line condition, the helper student was again provided with empty display cards and he (the helper student) provided the partially completed display cards (since his part of the assembly-line was to put one barrette on the card) to the target student.

Reliability

Reliability of the data was assessed for Woody and Fred during the Alone, Assembly-line, and return to Alone conditions. Reliability data were taken only in the Alone condition for Tim and Steve due to problems scheduling observers. Reliability consisted of the experimenter and one observer simultaneously and independently recording three measures on each task: 1) whether the student was working at the end of each 30 second interval, 2) number of completed products at end of each task (10 minute session), and 3) number of correctly completed products at the end of each task (10 minute session). The experimenter and observer sat at tables approximately five feet from the students when recording the data. At the end of each 10 minute session, each observer independently counted the total number of completed products and the number of correctly completed products. The measures of the students' on-task behavior were compared on an interval-by-interval basis. An agreement was scored if both observers recorded the student's behavior in the same way at the end of the interval. Reliability was calculated by dividing the number of agreements by the number of agreements plus disagreements and multiplying by 100. The measure of completed products (total number and correct) were compared by dividing the smaller score by the larger score and multiplying the quotient by 100.

RESULTS

Reliability

Reliability data for "on-task" behavior are summarized in Table 1.

Insert Table 1 about here

Mean reliability measures for Woody ranged from 85-100% across tasks/ conditions. They ranged from 80-99% for his partner, Fred. Tim's mean reliability measures ranged from 86-91% across tasks during the Alone condition; whereas Steve's means ranged from 97-99%.

Reliability data for the number of items completed ranged from 96-100% for Steve and were always at 100% for Tim. Measures on Woody ranged from 83-100% and those on Fred ranged from 78%-100%.

Number of Intervals Student was On-task

The results for each student are presented in graphic form. Figures 1, 2, 3, and 4, respectively, represent the percent of intervals in which Woody, Tim, Fred and Steve were on-task. High scores would be desirable on this measure. Each figure represents a single student's performance on the three different tasks used in the study. Each task is illustrated on a separate graph. The mean score for each condition is presented on each graph. Experimental conditions are separated by dashed lines.

Insert Figures 1, 2, 3, & 4 about here

Target Students. The performance of the severely handicapped target students is illustrated in Figures 1 and 2. In Figure 1, it can be seen Woody's performance is considerably higher in the Assembly-line condition than it is in either Alone condition on all three tasks. The

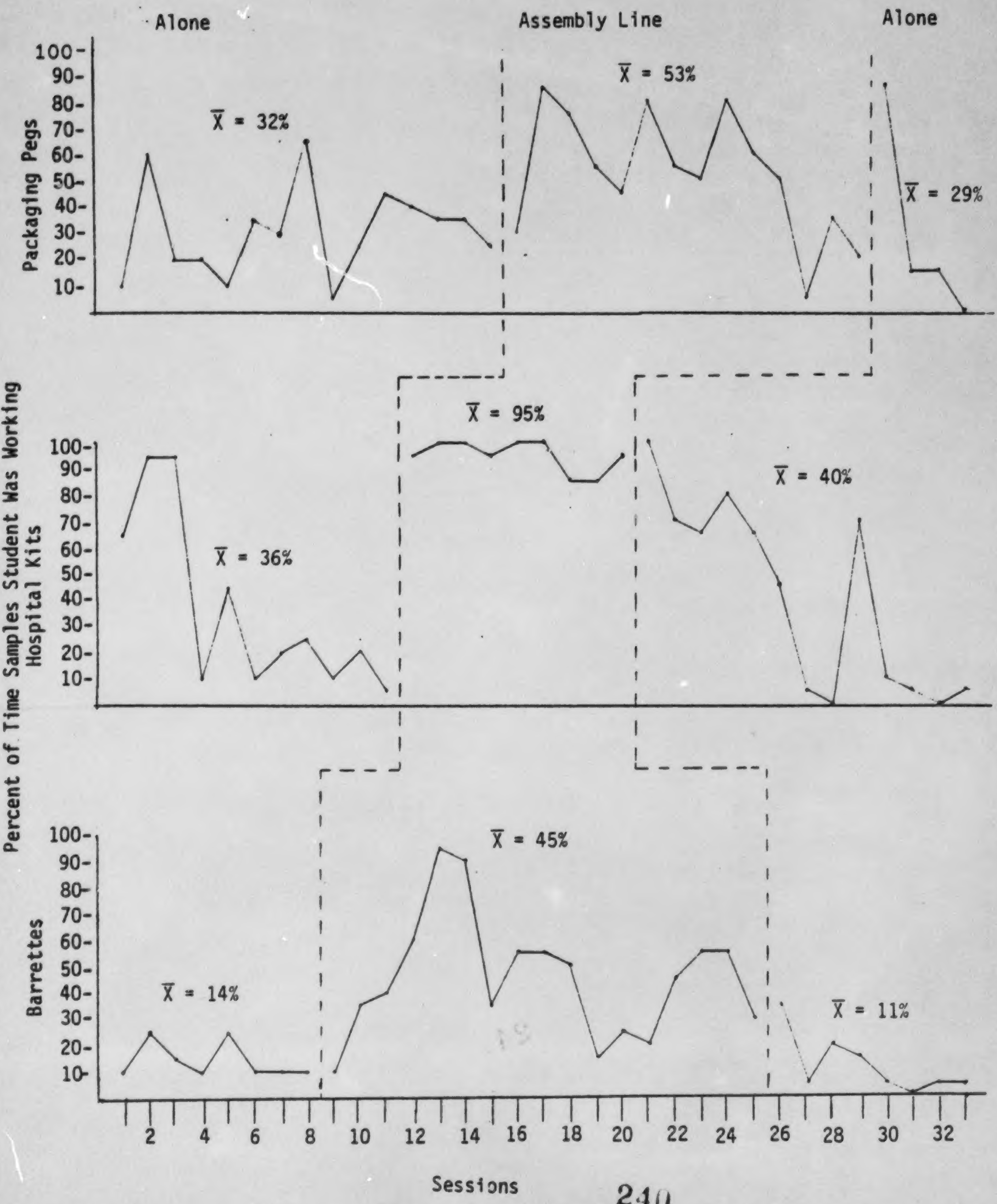
Table 1

Reliability of Time Sample Measures Across Tasks and Condition for Four Subjects

Subject	Task	Alone		Assembly-line	
		$\overline{X}$	range	$\overline{X}$	range
Woody	Hospital Kits	96%	89-100%	97%	95-100%
	Packaging Pegs	99%	95-100%	85%	70-100%
	Barretes	100%	100%	93%	75-100%
Tim	Hospital Kits	91%	85-100%	no data	-
	Packaging Pegs	86%	70-100%	no data	-
	Barrettes	91%	80-100%	no data	-
Fred	Hospital Kits	95%	85-100%	90%	80-95%
	Packaging Pegs	97%	90-100%	99%	95-100%
	Barrettes	80%	72-85%	93%	65-100%
Steve	Hospital Kits	99%	95-100%	no data	-
	Packaging Pegs	97%	92-100%	no data	-
	Barrettes	99%	94-100%	no data	-

Figure 1. Intervals of on-task behavior for Woody on three tasks under Alone and Assembly-line conditions.

Woody--SMH



Alone
Assembly Line
Alone
Percent of Time Samples Student Was Working
Packaging Pegs
$\overline{X}$ = 32%
$\overline{X}$ = 53%
$\overline{X}$ = 29%
Hospital Kits
$\overline{X}$ = 36%
$\overline{X}$ = 95%
$\overline{X}$ = 40%
Barrettes
$\overline{X}$ = 14%
$\overline{X}$ = 45%
$\overline{X}$ = 11%
100
90
80
70
60
50
40
30
20
10
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30
32
Sessions

Figure 2. Intervals of on-task behavior for Tim on three tasks under Alone and Assembly-line conditions.

Tim--SMH

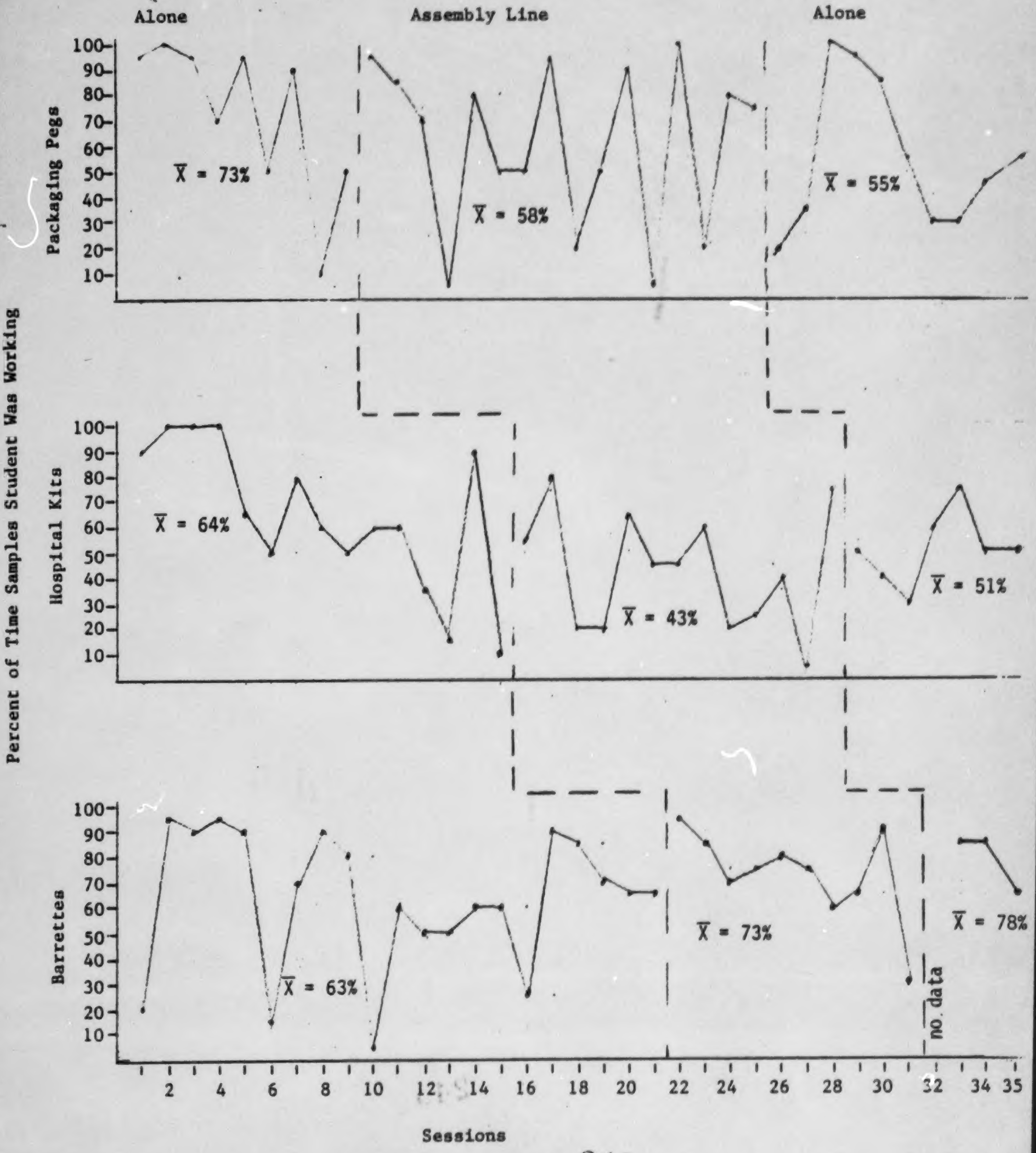

Alone
Assembly Line
Alone
Packaging Pegs
X̄ = 73%
X̄ = 58%
X̄ = 55%
Hospital Kits
X̄ = 64%
X̄ = 43%
X̄ = 51%
Barrettes
X̄ = 63%
X̄ = 73%
X̄ = 78%
no data
Percent of Time Samples Student Was Working
100
90
80
70
60
50
40
30
20
10
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30
32
34
35
Sessions

Figure 3. Intervals of on-task behavior for Fred on three tasks under Alone and Assembly-line conditions.

Fred--TMH

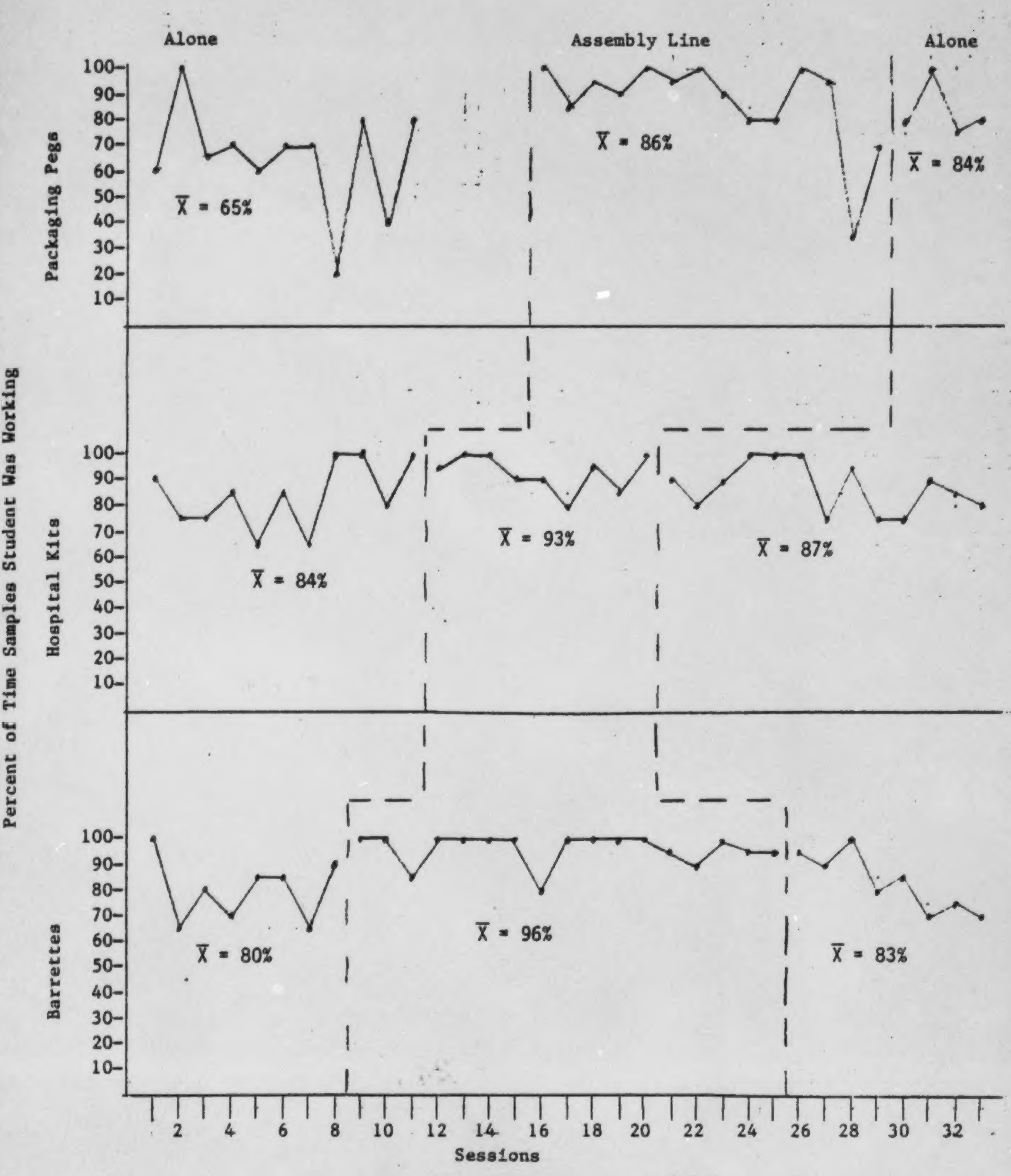
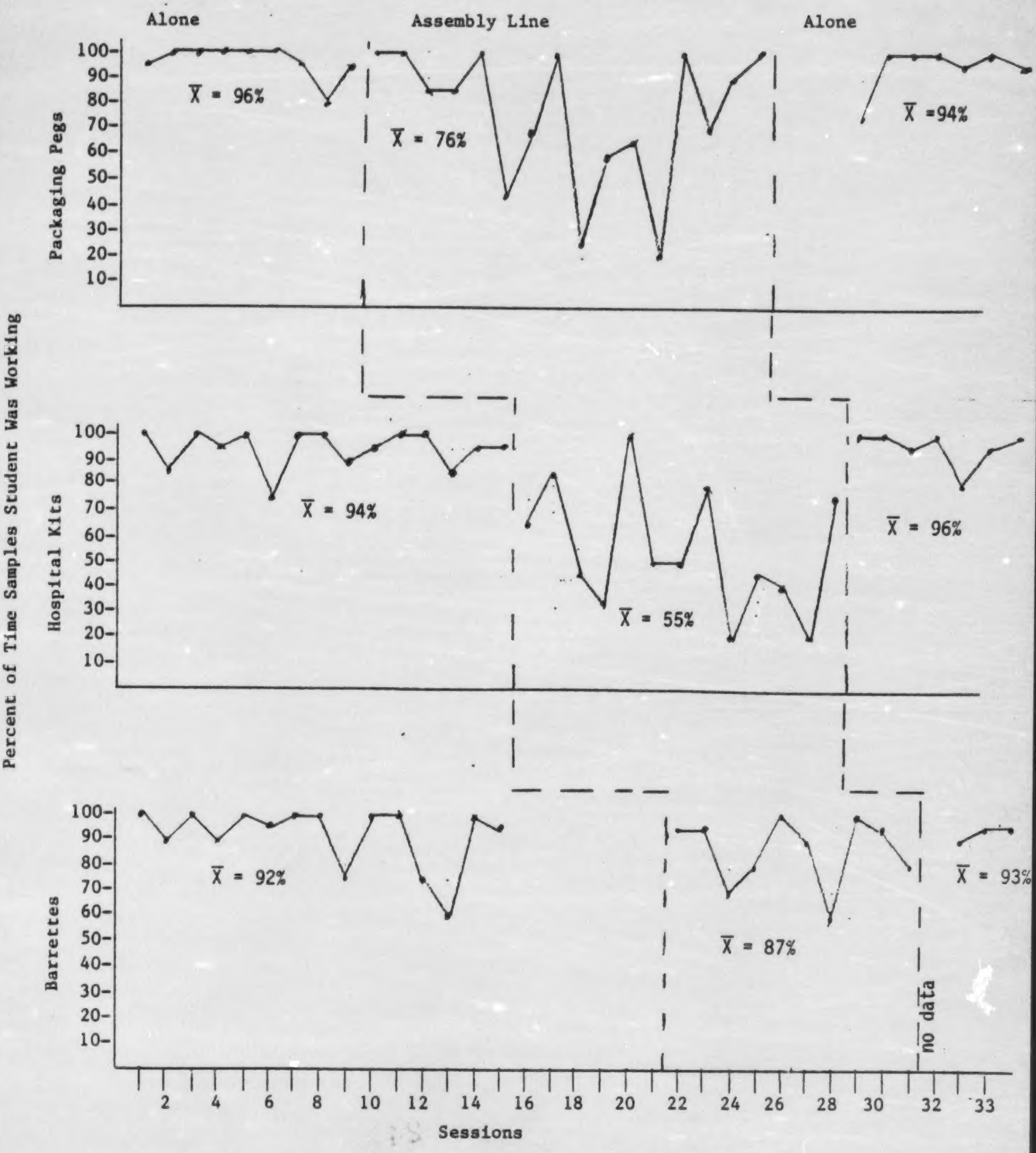

Alone
Assembly Line
Alone
Packaging Pegs
X̄ = 65%
X̄ = 86%
X̄ = 84%
Hospital Kits
X̄ = 84%
X̄ = 93%
X̄ = 87%
Barrettes
X̄ = 80%
X̄ = 96%
X̄ = 83%
Percent of Time Samples Student Was Working
10
20
30
40
50
60
70
80
90
100
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30
32
Sessions

Figure 4. Intervals of on-task behavior for Steve on three tasks under Alone and Assembly-line conditions.

Steve--TMH


Alone
Assembly Line
Alone
Percent of Time Samples Student Was Working
Packaging Pegs
X̄ = 96%
X̄ = 76%
X̄ =94%
Hospital Kits
X̄ = 94%
X̄ = 55%
X̄ = 96%
Barrettes
X̄ = 92%
X̄ = 87%
X̄ = 93%
no data
100
90
80
70
60
50
40
30
20
10
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30
32
33
Sessions

mean scores seen in the Assembly-line condition are almost double those seen in the Alone condition. There was considerable variability of performance across conditions, but the student seemed to achieve a stable performance under the Assembly-line condition for the Hospital kit assembly and under the Alone condition for the Barrette task.

In contrast, Tim's data in Figure 2 was higher in at least one of the Alone conditions for each of the three tasks. The differences in mean scores between conditions are not as large as those seen in Woody's data. Furthermore, very high variability in performance is seen both across, and within, conditions with this student.

Helper Students. The on-task performance of the two moderately retarded helper students is depicted in Figures 3 and 4. In Figure 3, it can be seen that Fred showed very high performance levels across conditions, but the mean scores indicate that Fred, like his partner Woody, spent more time working during the Assembly-line condition.

Figure 4 shows that Steve consistently was on-task under the Alone condition, but his performance was considerably lower and more variable under the Assembly-line condition.

Number of Items Completed/Correctly Completed

Figures 5-8 illustrate both the total number of items completed and the number of correct items completed by Woody, Tim, Fred, and Steve,

Insert Figures 5, 6, 7, & 8 about here

respectively. The number of items completed during a session is represented by a solid circle whereas the number of items correctly completed is represented by an X. If only an X is marked for a session, this indicates that all the items completed during this session were correct.

Figure 5. Number of correct items and items completed by Woody on three tasks under Alone and Assembly-line conditions.

Woody--SMH

· = items completed
x = items correctly completed

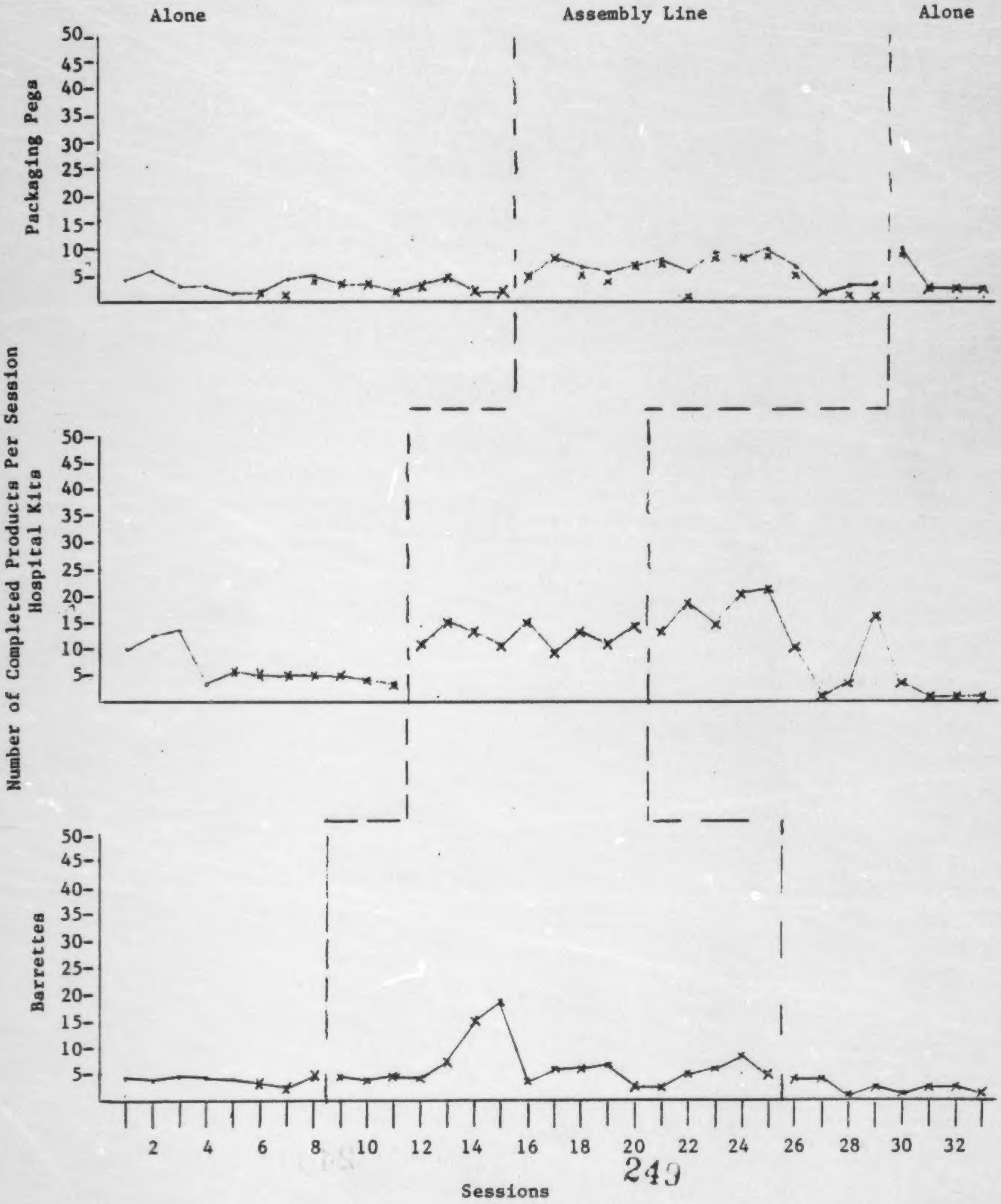


Figure 6. Number of correct items and items completed by Tim on three tasks under Alone and Assembly-line conditions.

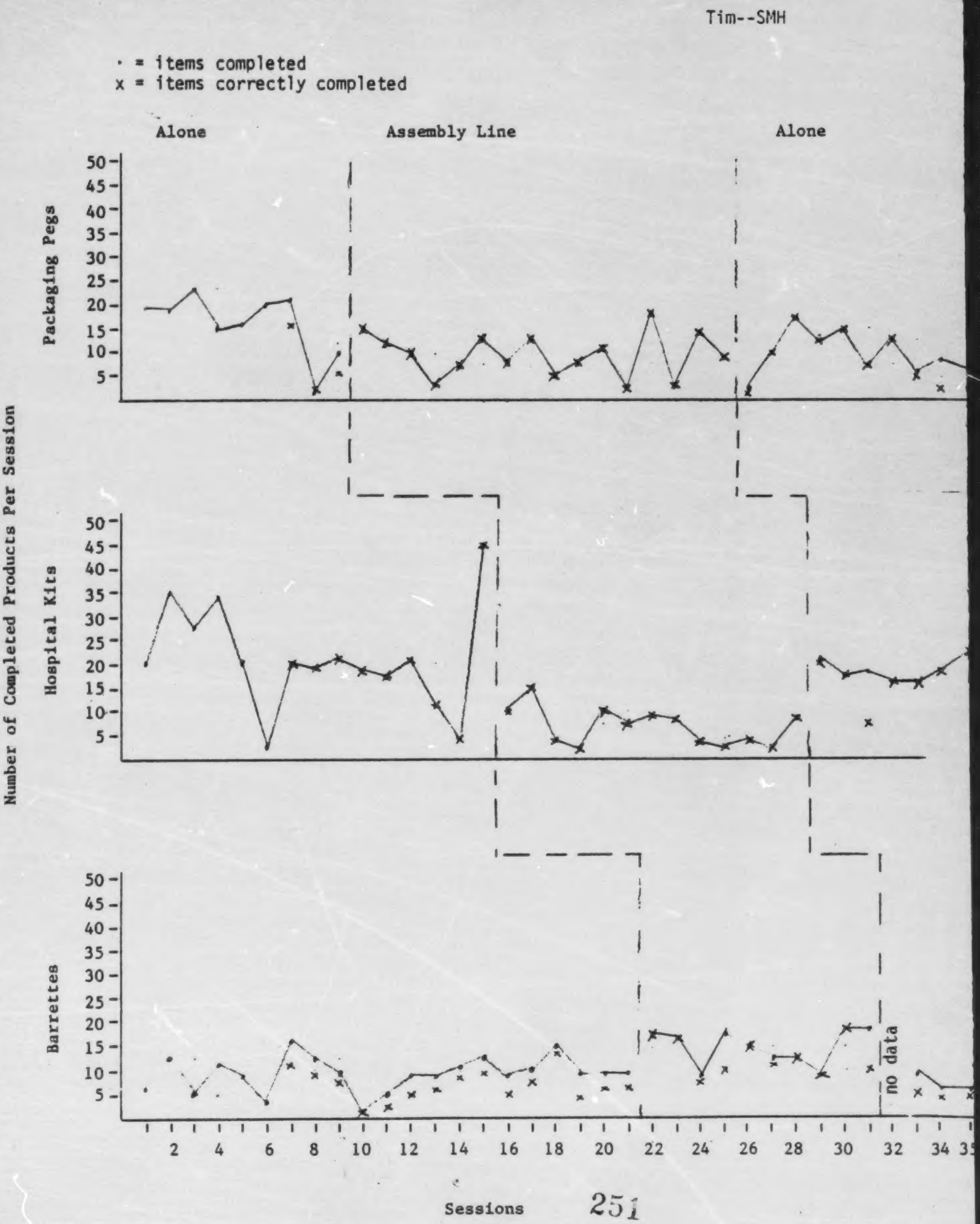

Tim--SMH
· = items completed
x = items correctly completed
Alone
Assembly Line
Alone
Number of Completed Products Per Session
Packaging Pegs
Hospital Kits
Barrettes
50
45
40
35
30
25
20
15
10
5
no data
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30
32
34
Sessions

Figure 7. Number of correct items and items completed by Fred on three tasks under Alone and Assembly-line conditions.

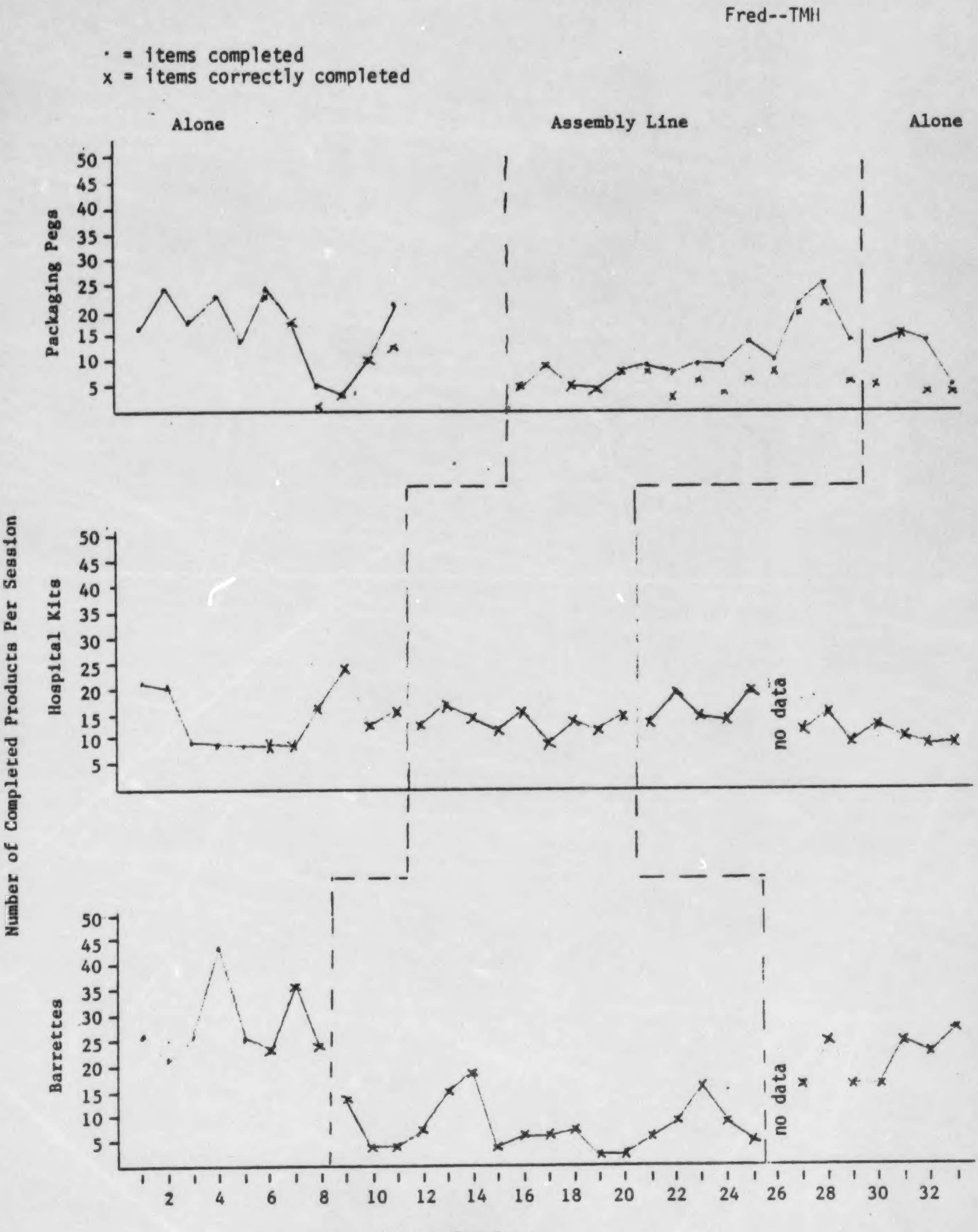
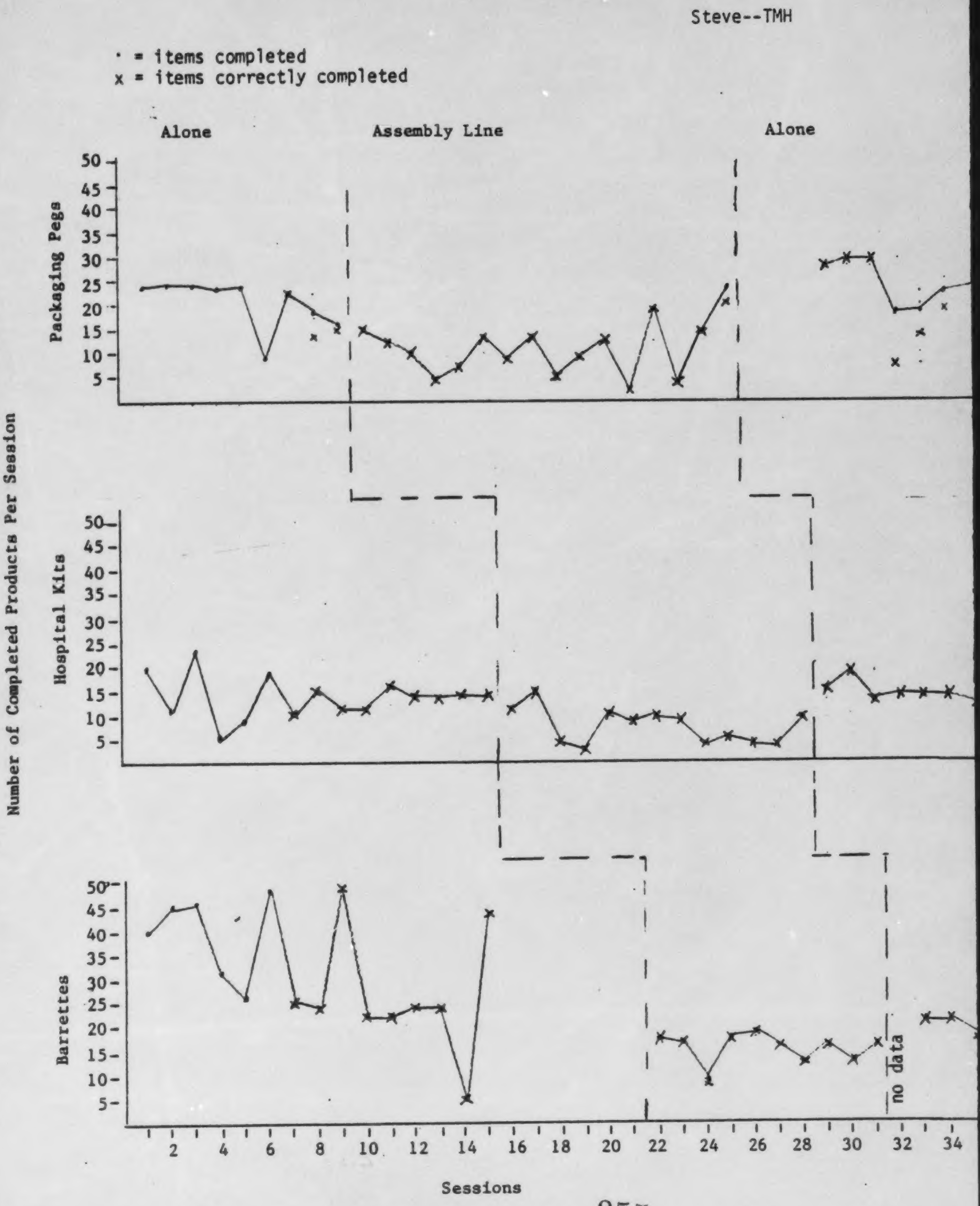

Fred--TMH
· = items completed
x = items correctly completed
Alone
Assembly Line
Alone
Number of Completed Products Per Session
Packaging Pegs
Hospital Kits
Barrettes
no data
no data
50
45
40
35
30
25
20
15
10
5
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30
32
Sessions

Figure 8. Number of correct items and items completed by Steve on three tasks under Alone and Assembly-line conditions.

Steve--TMH
· = items completed
x = items correctly completed
Alone
Assembly Line
Alone
Packaging Pegs
Hospital Kits
Barrettes
Number of Completed Products Per Session
50
45
40
35
30
25
20
15
10
5
no data
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30
32
34
Sessions

Target Students. Woody (Figure 5) consistently completed slightly more items in the Assembly-line condition than he did in the Alone condition on all three programs, though there was a brief increase in Hospital Kit assembly at the beginning of the second Alone condition. Variability was generally low within and across conditions on a task. Although this student had a low level of production in the three tasks, he appeared to do his work accurately. There were no incorrectly completed products in either the Hospital Kit or Barrette task, regardless of condition. On the Packaging Pegs task, it appears that the student made more errors during the Assembly-line condition than in either Alone condition. The mean scores shown in Table 2, however, show that Woody not only completed more items in the Assembly-line condition, but also completed more items correctly.

Tim (Figure 6) completed more items during both the Alone conditions than he did in the Assembly-line condition on the first two tasks. On the other task (Barrettes), he completed more in the Assembly-line condition than he did in either Alone condition. The number of items completed and those completed correctly were quite variable from session to session. A comparison of Figures 2 and 6 shows one interesting fact. On the Barrette task, Tim got the fewest number of items completed in the second Alone condition, in spite of the fact that he spent more time on-task in that condition than he had in the other two conditions. Tim, like Woody, showed a high degree of accuracy on two of the tasks. On the Barrette task, however, his accuracy was poor in both conditions. Tim's data in Table 2 indicate that more items were correctly completed

----

Insert Table 2 about here

Table 2

Mean Number of Items Correctly Completed and Completed During Alone and Assembly-Line Conditions

Student	Task	Alone		Assembly-line		Alone	
		Complete	Correct	Complete	Correct	Complete	Correct
Woody	Packaging Pegs	3.2	2.5	6.2	5.1	4.0	3.8
	Hospital Kits	7.7	4.7	12.3	12.3	9.4	9.4
	Barrettes	3.4	3.0	6.2	6.2	2.1	2.1
Fred	Packaging Pegs	15.6	11.0	10.5	7.6	11.5	6.5
	Hospital Kits	13.5	12.9	12.6	12.6	12.4	12.4
	Barrettes	28.0	27.0	7.7	7.7	21.1	21.1
Tim	Packaging	16.0	8.0	9.4	9.4	9.8	8.4
	Hospital Kits	20.9	19.6	6.5	6.4	18.3	16.6
	Barrettes	9.2	6.6	13.9	12.1	6.7	4.7
Steve	Packaging Pegs	20.2	16.7	10.6	10.4	21.6	18.6
	Hospital Kits	13.6	13.1	6.8	6.8	13.9	13.9
	Barretes	31.5	26.2	14.8	14.7	20.0	19.7

in the Alone condition when Packaging pegs and Barrettes, but the Hospital Kits were more accurately done during the first Alone condition.

Helper Students. Both Fred and Steve were consistently able to complete more items during the Alone conditions than they were under the Assembly-line condition, though the differences were more marked in Steve's case. Variability was high across sessions for both students. Steve was generally accurate in his performance, though he had several sessions of inaccurate performance in the two Alone conditions when Packaging pegs. Fred made no errors when doing the Hospital Kit or the Barrettes. He had considerable difficulty in packaging pegs correctly, however, and this difficulty was seen in all conditions. The mean scores shown in Table 2 show that Steve's productivity and accuracy were enhanced in the two Alone conditions. Fred, on the other hand, showed his highest productivity and accuracy in the first Alone condition, but this trend did not always occur in the second Alone condition.

DISCUSSION

The data for Woody confirmed the hypothesis that the production ratio and on-task behavior of a student, who manifested production problems due to slow motor behavior, could be improved by an assembly-line format. There was no loss in accuracy with this increase in production. Thus, it would appear that changing the work format could have beneficial effects on those students with severe handicaps who have learned how to do a task but need to be constantly reminded to stay on task. This change in format might allow a student who is dependent on prompts to be successful in the workplace. It should, however, be pointed out that increasing the target student's production was, to some extent, done at the expense of the less handicapped student's production. Fred, who was Woody's partner, also showed more on-task behavior during the Assembly-line condition, but completed less items and made more errors during this condition. This drop in production and accuracy may have occurred because Fred was attending to the target student and attempting to keep him on-task, resulting in less attention to his own work.

The results obtained with Tim did not confirm the hypothesis that low production rates that seem to be correlated with interfering behavior can be increased by using an assembly-line format. Overall, Tim spent more time on-task, produced more items and was more accurate during the Alone condition, though there was a slight increase in production on the Barrette task under Assembly-line conditions. Tim's partner, Steve, similarly showed much higher on-task behavior, production and accuracy during the Alone condition. This superiority of the Alone condition for these students appeared to be due to the types of interaction that occurred between the students. The helper student gave the target student several verbal cues to get busy during the sessions, and at times threatened to "fire the helper student if he didn't straighten up." When the target student was working, however, the helper did not interact with him other than to give him more materials. The target student seemed to be highly reinforced by the interactions with the helper in spite of their apparent negative content, and the observers all commented on the fact that Tim began to make more errors and to dawdle a great deal more after such an interaction. These negative interactions also took a great deal of the helper student's time, and probably accounted for the losses seen in the helper's production during the Assembly-line condition. It would be interesting to determine whether such an effect would have been found if the helper student had been instructed to interact with the target student only when he was working appropriately.

The results in this study appear to confirm those in a similar study conducted by Brown, Johnson, Gadberry and Fenrick (1971) who found that the production rates of moderately handicapped students were consistently higher under individual work conditions than under assembly-line conditions. Their study differed from the present study in that they changed the number of different behaviors in which the student had to engage when the student was in the assembly line condition. In spite of this major difference between the studies, the data in the current study also show the superiority of individual work for the students who were moderately handicapped.

Though it is possible that these results are unique to the students in this study, the success of the change in format for the student with severe handicaps who had slow motor behavior was quite encouraging and may be useful for those practitioners who are struggling with a similar problem. Since this improvement in one student's production appeared to be tied to a decrement in the more able student's production, however, it may not be an acceptable solution in the workplace. It would, thus, be useful to determine whether a limited training period in an assembly-line format would result in durable effects if the target student was slowly phased into individual work.

References

Bellamy, G. T., Peterson, L., and Close, D. Habilitation of the severely and profoundly retarded: Illustrations of competence. Education and Training of the Mentally Retarded, 1975, 10, 174-186.

Bellamy, G. T., Sheehan, M., Horner, R., & Boles, S. Community programs for severely handicapped adults: An analysis. Journal of the Association for the Severely Handicapped, 1980, 5, 307-324.

Brown, L., & Pearce, E. Increasing the production rate of trainable retarded students in a public school simulated workshop. Education and Training of the Mentally Retarded, 1970, 5, 15-22.

Brown, L., Johnson, S., Gadberry, E., & Fenrick, N. Increasing individual and assembly-line production rates of retarded students. Training School Bulletin, 1971, 67, 206-213.

Brown, L. F., Van Deventer, P., Perlmutter, L., Jones, S., & Sontag, E. Effects of consequences on production rates of trainable retarded and severely emotionally disturbed students in a public school workshop. Education and Training of the Mentally Retarded, 1972, 7, 74-81.

Clarke, J., Greenwood, L., Abramovitz, D., & Bellamy, G.T. Summer jobs for vocational preparation of moderately and severely retarded adolescents. Journal of the Association for the Severely Handicapped, 1980, 5, 24-37.

Gold, M.W. Stimulus factors in skill training of the retarded on a complex assembly task: Acquisition, transfer, and retention. American Journal of Mental Deficiency, 1972, 76, 517-526.

Gold, M. W. Task analysis: A statement and an example using acquisition and production of a complex assembly task by the retarded blind. Exceptional Children, 1976, 43, 78-84.

Gold, M.W., and Barclay, C. R. The learning of difficult visual discrimination by the moderately and severely retarded. Mental Retardation, 1973, 11, 9-11.

Hersen, M. & Barlow, D. H. Single-case experimental designs: Strategies for studying behavior change. New York: Pergamon Press, 1976.

Irvin, L. K. General utility of easy-to-hard discrimination training procedures with the severely retarded. Education and Training of the Mental Retarded, 1976, 11, 247-250.

Kliebhahn, I. J. Effects of goal setting and modeling on job performances of retarded adolescents. American Journal of Mental Deficiency, 1967, 72, 220-226.

Levy, S. M. The development of work skill training procedures for the assembly of printed circuit boards by the severely handicapped. AAESPH Review, 1975, 1(1).

O'Neill, C., & Bellamy, G. T. Evaluation of a procedure for teaching saw chain assembly to a severely retarded woman, Mental Retardation, 1978, 16, 37-41.

Renzaglia, A., Wehman, P., Schutz, R., & Karan, O. C. Use of cue redundancy and positive reinforcement to accelerate production in two profoundly retarded workers. In O. C. Karan, P. Wehman, A. Renzaglia, & R. Schutz (Eds.), Habilitation practices with the severely developmentally disabled, Volume 1. Madison, WS: University of Wisconsin, Department of Special Educaiton, 1976.

Revell, W. G., Arnold, S., Taylor, B., & Zaitz-Blotner, S. Project transitions: Competitive employment service. Journal of Rehabilitation, 1982, 48, 31-35.

Rusch, F. R., Connis, R., & Sowers, J. The modification and maintenance of time spent attending to task using social reinforcement, token reinforcement and response cost in an applied restaurant setting. Journal of Special Education Technology, 1979, 2, 18-26.

Sailor, W. & Mix B. J. The TARC assessment system. Lawrence, KS: H and H Enterpreses, 1975.

Schroeder, S. Parameteric effects of reinforcement frequency, amount of reinforcement, and required response force on sheltered workshop behavior. Journal of Applied Behavior Analysis, 1972, 5, 431-441.

Spooner, F. & Hendrickson, B. Acquisition of complex assembly skills through the use of systematic training procedures: Involving profoundly retarded adults. AAESPH Review, 1976, 1(7), 14-25.

Wehman, P. and Hill, B. (Eds.) Vocational training placement of severely disabled persons. Richmond, VA: Virginia Commonwealth University, School of Education, 1982.

Wehman, P. & McLaughlin, P. J. *Vocational curriculum for developmentally disabled persons*. Baltimore, MD: University Park Press, 1980.

List of Figures

1. Intervals of on-task behavior for Woody on three tasks under Alone and Assembly-line conditions.

2. Intervals of on-task behavior for Tim on three tasks under Alone and Assembly-line conditions.

3. Intervals of on-task behavior for Fred on three tasks under Alone and Assembly-line conditions.

4. Intervals of on-task behavior for Steve on three tasks under Alone and Assembly-line conditions.

5. Number of correct items and items completed by Woody on three tasks under Alone and Assembly-line conditions.

6. Number of correct items and items completed by Tim on three tasks under Alone and Assembly-line conditions.

7. Number of correct items and items completed by Fred on three tasks under Alone and Assembly-line conditions.

8. Number of correct items and items completed by Steve on three tasks under Alone and Assembly-line conditions.

Evaluation of Data Recording Systems

The studies contained in this section were requested by classroom staff who were interested in reducing their paperwork burden. The studies looked at the issues of how often data need to be collected, whether data need to be collected regularly on an educational program which has been taught for several months under a data base, and whether the format of the data sheet has a direct impact on the precision of teaching.

Variables of interest in these studies focused on: 1) the precision with which a least prompts strategy was used; 2) the precision with which a teacher followed a task-analysis; 3) determination of learning trends; 4) knowledge of how well the student did in each session; and 5) educational decision-making.

These studies supported the need for frequent data recording, especially when educational decisions need to be made. On the other hand, the studies also indicated that it may not be necessary to collect data on a daily basis. The format of the data sheet, apparently, has a minimal impact on the precision with which the least prompts strategy is used.

How Many Data Are Enough?

Jennifer F. Holvoet, Ph.D.

Marilyn Mulligan, Ph.D.

Doug Guess, Ed.D.

Nancy Schussler, M.S. Ed.

Department of Special Education
University of Kansas

1983

This paper was prepared as part of the work pursuant to the "Kansas Individualized Curriculum Sequencing Project" funded as a Handicapped Children's Model Project, U.S. Office of Education (G00-800-1721) from 9-1-80 to 9-1-83.

Abstract

Fourteen teachers compared daily versus less-than-daily performance graphs of severely handicapped students fo purposes of interpreting progress and making instructional decisions. Results were analyzed according to intrasubject reliability and with respect to the number of data points plotted on graphs (i.e., daily versus less-than-daily). It was found that teachers were reliable in interpreting the direction of graphs at two separate points in time, but were inconsistent in making decisions for program changes. Results also showed no differences in the interpretation of directional changes of daily versus less-than-daily graphs; but some differences were found in making program decisions on the number of data points presented.

Data collection is seen by most professionals in the field of severely handicapped education as one of the most valuable tools available to teachers (c.f., Hanson, 1978; Sailor & Haring, 1977; Wehman, Bates, & Renzaglia, 1980). Through data collection a teacher can: 1) assess a student's level of functioning before implementing instruction; 2) monitor the student's progress toward an educational objective; 3) change instructional techniques when data analysis shows that learning is not taking place; and 4) discuss a student's progress in an objective manner with parents and other interested parties.

Because data can serve all these important functions, many professionals have put a great deal of emphasis on teaching prospective teachers of severely handicapped students a variety of data recording systems. This includes some fairly quick methods of interpreting the data and making data-based instructional decisions. Unfortunately, one of the important parameters that generally is not adequately addressed is how often data need to be collected. Professional articles about optimal teaching practices (Browning, 1980; Fredricks, Anderson, & Baldwin, 1979; Gentry & Haring, 1976; Mori & Masters, 1980; Wehman, Bates, & Renzaglia, 1980) and observations of classrooms for severely handicapped students indicate teachers are encouraged to collect instructional data on a daily basis. Many teachers, on the other hand, seem to agree with Edgar (1978) who observed,

> "It is important to keep in mind that you can collect some of the data all of the time, or all of the data some of the time; but you can't collect all of the data all of the time" (p. 37).

Thus, it appears that there is a need to investigate how frequently data should be taken in order to assist teachers in making sound instructional decisions for severely handicapped students.

This study looked at the question of whether less-than-daily data are comparable to daily data for the purposes of: 1) deciding how a student is progressing (Interpretation), and 2) making instructional decisions (Decision-making).

Method

Subjects/Setting

Fourteen individuals served as subjects for this study. Eleven of these individuals were enrolled in a graduate level special education course focusing on methods of teaching severely handicapped students. This group was composed of teachers of the severely handicapped. The other three subjects were physical and occupational therapists attending a similar class, as part of a therapist-consultant training program. Two different instructors taught the two groups of subjects. Most of the subjects (regardless of campus) had spent at least one year working in classrooms serving severely multiply handicapped students. These individuals had all been taught to record daily data when doing instructional tasks with their students and to make instructional decisions from these data. This training had occurred during the previous semester with the same instructor they had during the study.

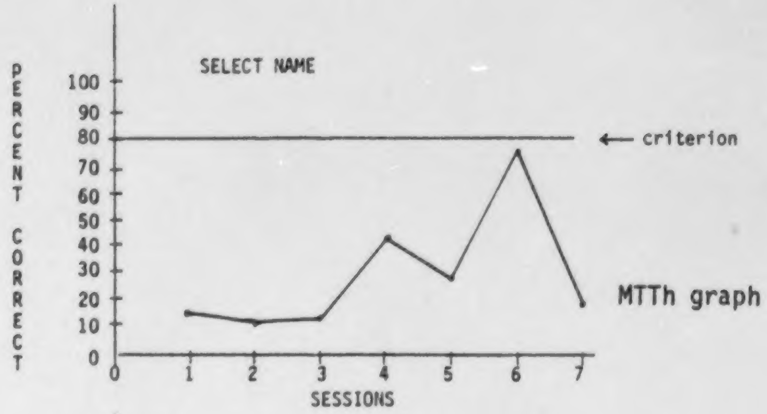
Materials

The graphs used for decision-making were drawn from actual data collected on eight programs from each of two severely handicapped students (a total of 16 different programs). One of these students was observed by classroom staff to be a relatively quick learner and the other was characterized as a relatively slow learner. Three weeks of data were chosen for analysis. Four graphs were drawn from the data on each of the 16 programs: (1) a graph which had the daily scores plotted on it (Daily); (2) a graph which had the scores from only Monday and Wednesdays plotted on it (MW); (3) a graph which had the scores from only Tuesdays and Thursdays plotted (TTh); and (4) a graph which had the scores from Monday, Tuesday and Thursday on it (MTTh). These four versions of the program data constituted the independent variables in this study. If the student had been absent, or the program was not taught on a given day (e.g., a Monday), no data point was plotted. The minimum number of points on a graph was 4 and the maximum was 15. In addition to the data points, a graph included the student's name (fictitous), the name of the program, session numbers, a labeled ordinate, and a criterion line. All graphs were equal-interval graphs. The graphs did not contain any information as to whether they were Daily, MW, TTh, etc. (except for the fact that some had fewer data points). Figure 1 is an example of the four graphic versions of the data for one of the 16 programs included in the study.

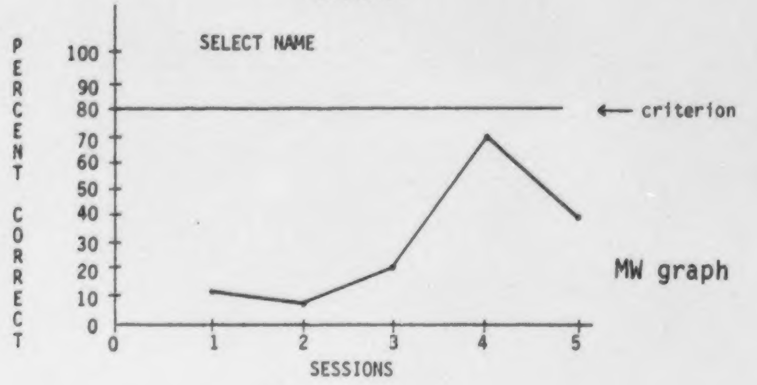
Insert Figure 1 about here

The graphs were arranged in packets when given to the subjects. A packet consisted of one graph for each program (16 graphs total). Daily, MW, TTh, and MTTh graphs were included in each packet. Each of the four graphs from a program was randomly assigned to a different packet; therefore, four different packets were made. These four packets were labeled A, B, C, and D. In order to assess intrasubject reliability, the A packet was reproduced a second time and was called packet E. The makeup of the graphs comprising each packet is delineated in Table 1.

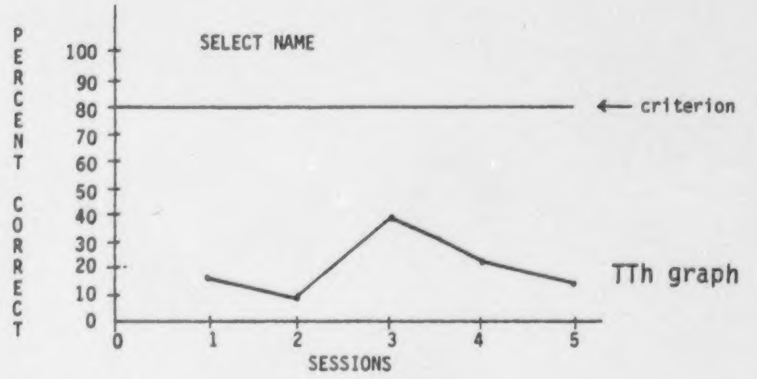
Figure 1. An example of the four graphic formats used for analysis.



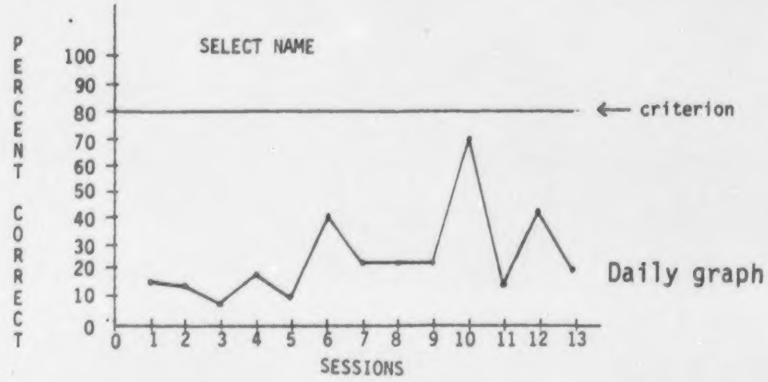
MTTh graph



MW graph



TTh graph



Daily graph

Insert Table 1 about here

Included with each packet was a questionnaire. The questionnaire stated three questions to be answered for every graph (program). These questions were as follows:

1) These data are basically: a) going up; b) going down; c) staying the same below criterion; and d) above criterion for three or more sessions.
2) The program should: a) continue as is; b) be ended; c) be changed; and d) be watched carefully.
3) If you marked alternative 2c above, which change should be made? a) change reinforcer; or b) change step to something easier or use a different method to teach the same task.

The subjects' answers to these three questions comprised the dependent variables of this study.

Procedures

The first packet was administered during the third week of class in the semester with new packets being given approximately every two weeks thereafter. The packets were distributed and completed during class time. The packets were presented in the following order for the eleven Special Education teachers: B, A, C, D, E. The packets were presented in order from A to E for the three therapists.

Subjects who were absent during the class period in which a particular packet was to be completed were given that packet to complete at home. Two Special Education teachers took two packets home at one week intervals.

Subjects were verbally instructed to: 1) read the questionnaire; 2) locate the designated graph (i.e., be sure they were looking at Graph

Table 1

The Graphic Format Used for the Packets in Each of 16 programs

PROGRAM	PACKET			
	A/E	B	C	D
1. Sign "Money"	TTh	MTTh	Daily	MW
2. Sign "Cup"	TTh	Daily	MW	MTTh
3. Sit down	MW	Daily	TTh	MTTh
4. Select name	MTTh	MW	TTh	Daily
5. Picture/Object match	Daily	MTTh	MW	TTh
6. Touch quadrants	Daily	MW	TTh	MTTh
7. Come here	MTTh	MW	TTh	Daily
8. Sign "Hi"	TTh	Daily	MW	MTTh
9. Use of identification card	MTTh	MW	Daily	TTh
10. Drinking correctly	MW	TTh	MTTh	Daily
11. Unzipping	Daily	MTTh	MW	TTh
12. "What want?"	MW	TTh	Daily	MTTh
13. Vacuuming	Daily	TTh	MW	MTTh
14. Receptive labeling	MW	TTh	MTTh	Daily
15. Empty trash	Daily	TTh	MTTh	MW
16. Zipping	MTTh	Daily	TTh	MW

#7 when answering the questions about Graph #7; 3) decide which of the multiple choice statements best described the data and the decisions they would make based on the graph; and 4) mark their choice by circling the desired answer. They were instructed to be sure to complete all portions of the questionnaire and not to leave out any questions.

The students typically asked two questions when completing the questionnaire: 1) Is this an acceleration or deceleration program? and 2) How many days of data are represented? The instructor answered these questions, respectively: 1) "The criterion line is indicated." and 2) "Three weeks of data are represented on all the graphs." Other questions were addressed by the answer, "Look at the graph and make your decisions based on the information given."

Data Analyses

Questionnaire data were tabulated so that responses from each subject could be compared within and across programs represented by the Daily, MW, TTh, and MTTh graphs. Two levels of analyses were performed on the data. The first level analyzed intrasubject reliability across the following three questions (dependent variables):

1) Do the teachers interpret the graphs the same way when they look at the same graphs at two separate points in time?;
2) Do the teachers make the same programming decisions when they look at the same graphs at two separate points in time?;
3) If a teacher thought a change should be made in the program, was the same change made when the same graphs were analyzed at a later time?

The intrasubject reliability measures were made by comparing the identical graph packets, A and E. This type of analysis was necessary before the second level of analysis could be undertaken.

The second level of analysis was a comparison between the Daily and the MW, TTh, and MTTh graphs across each of the three questions that constituted the dependent variables (i.e., the same three questions used in analyzing intrasubject reliability).

Comparisons of MW to MTTh data, TTh to MW Data, or TTh to MTTh data were not made because the question of interest centered around whether teachers would interpret and make the same decisions about a program if they recorded data intermittently rather than on a daily basis.

Data from the two subject groups were combined after preliminary analyses showed no differences between the obtained responses. All measures were calculated using the following formula for a one-tailed binomial test (Conover, 1971):

$$t_2 = N(p) + (k)\sqrt{N\,(p)(1-p)}$$

The selected confidence level was .05 for all measures.

Results

Intrasubject Reliability

Using the binomial test, the subjects' responses to the graphs contained in packet A were compared to their responses to the graphs contained in packet E. These results are summarized in Table 2.

Insert Table 2 about here

For question 1 (Do the teachers interpret the graphs the same way when they look at the same graphs at two separate points in time), the data indicated that the teachers were reliable on all formats (see Table 2, Column 1).

Table 2

Intrasubject Reliability Scores Across Questions on the Various Grahpic formats Contained in Packets A and E Calculated with a One-Tailed Binomial Test

Format	Question 1		Question 2		Question 3	
	Rejection value[1]	Obtained value	Rejection value	Obtained value	Rejection value	Obtained value
All formats combined	92.2	43	92.2	64	24.5	19
MTTh	18.3	12	18.3	22*	5.5	4
MW	20.1	8	20.1	10	9.2	7
TTh	15.7	12	15.7	16*	-	-
Daily	26.3	11	26.3	17	5.1	7*

[1] If the rejection value is derived from the binomial distribution table is larger than the obtained value, then the scores obtained on packets A & E were not significantly different.

* Difference is significant at .05 level.

For question 2 (Do the teachers make the same decisions when they look at the same graphs at two separate points in time), the data (see Table 2, Column 2), indicated that the teachers were reliable on the MW format and on the Daily format. They, however, did not reliably make the same decisions when using the TTh or the MTTh formats.

For Question 3 (If a teacher thought a change should be made in the program, was the same change made when analyzing the same graphs at a later time), a comparison was made only on those programs where, on Question 2, the subject had said a change was needed on both sets (A & E) of graphs. This was done to prevent carrying over error from Question 2 to Question 3. The data (see Table 2, Column 3) indicated that the teachers were reliable on the MW and the MTTh graphs. They, however, were not reliable on the Daily data for this question. No data were available for the TTh graphs, because there were no TTh graphs in Sets A/E that the teachers felt needed a program change.

Comparisons of Daily with MW, TTh, and MTTh Data

Question 1 Interpretation. Question 1 addresses the issue of whether teachers will interpret data that are taken less frequently in the same way (i.e., as going down, at criterion, etc.) as they interpret daily data. The teachers' answers to Question 1 on the Daily graphs were compared to their answers on the MW graphs for all programs. If either the Daily or the MW data fell in packets A/E, the data from the E set were used for comparison. These data are summarized in Table 3. A binomial test with confidence levels set at .05 showed that Daily and MW data were seen as comparable.

Insert Table 3 about here

Table 3

Comparison of Daily vs. MTTh, MW, and TTh Graphic Formats for Each of the Three Questions Calculated with a One-Tailed Binomial Test

Comparisons	Question 1		Question 2		Question 3	
	Rejection value[1]	Obtained value	Rejection value	Obtained value	Rejection value	Obtained value
Daily vs MTTh	94.0	65	-	-	-	-
Daily vs MW	94.5	79	94.5	103*	-	-
Daily vs TTh	90.0	83	-	-	-	-

[1] If the rejection value is larger than the obtained value, then the scores obtained from the compared graphic formats were not significantly different.

* Difference is significant at .05 level.

The same method was used to compare interpretations of Daily to TTh data, and Daily to MTTh data. Both these comparisons showed that teachers interpreted the less frequent data in the same way that they interpreted the daily data.

Question 2: Instructional decisions. The second question addressed the issue of whether teachers, given less than daily data, make the same instructional decisions they would make when given daily data. The answer each teacher gave to Question 2 on the Daily graph was compared to the answer given to Question 2 on the MW graph for each program. If either the Daily or the MW data were included in the A/E packets, the data from the E set were used for comparison. A binomial test with the confidence levels set at .05 was used to assess the degree of agreement. These data are summarized in Column 2 of Table 3. These data indicate that teachers do not make the same instructional decisions when given MW data as they do when given Daily data.

Comparisons of Daily to TTh and Daily to MTTh decisions could not be made because of the lack of intraobserver reliability on the TTh and MTTh data for Question 2.

Question 3: Recommending changes. Question 3 looked at the issue of whether a teacher who, looking at daily data, thought a change needed to be made in a program would make the same change when given less data. Unfortunately, this comparison could not be made because the Daily data on Question 3 were not reliable within subjects.

Discussion

The first concern of this study was whether teachers would reliably assess and analyze the same graphs when viewed at separate periods in time, regardless of the number of data points available to them. Results showed that, as a group, the teachers interpreted the graphs in the same manner, independent of the number of data points presented on the graphs (Question 1). They were, in effect, reliable in their perceptions as to whether the data points were going up, down, or staying the same; and in determining whether performance, was at criterion levels.

For Question 2, the teachers agreed with themselves on program changes (or the decision not to change) on both the Daily and MW graphs. As a group, however, their recommendations for program changes were not reliable for either the TTh or MTTh graphs. This finding is somewhat perplexing because the number of data points for the MW graphs (maximum of six points) was essentially the same as for the TTh graphs, and even less than for the MTTh graphs (maximum of nine points).

Intrasubject reliability for Question 3 analyzed whether teachers were consistent in their recommendations for the type of change to be made in a program; e.g., a) change reinforcer; or b) change step to something easier or use a different instructional method.

These data were analyzed only for those programs where the teachers had indicated the need for change (Question 2) on both sets of graphs, A and E. Results showed they were reliable on the MWTh and the MTTh graphs, but not on the Daily graphs. As mentioned earlier, data were not available on the TTh graphs because teachers did not perceive the need for program changes.

Overall, the intrasubject reliability measures showed that the teachers were consistent in their ability to read the directions of the graphs, regardless of the number of data points available to them. They were, however, inconsistent in their recommendations for program changes for both TTh and MTTh graphs; and they were not reliable on the type of change selected for data included in the Daily graphs. This latter finding is especially disturbing because making the appropriate (and consistent) procedural change is one of the major purposes for collecting data in the first place.

The second issue of concern in this study was how teachers' analyzed daily versus less-than-daily data. Results from Question 1 showed that there were no significant differences between interpretations made on daily data and those made on data taken twice a week or three times a week. This finding supported some teachers' contention that they could accurately determine how well a student is learning, even if they do not take daily data. Results also indicated that teachers could take data as few as two times a week without losing track of the students' performance.

Unfortunately, both the second and third questions of the study were affected, adversely, by the lack of intrasubject reliability. When analyzing the question (2) of the instructional decision process only the Daily versus MW data could be compared. This comparison, indicated significant differences between decisions made on these separate data sets. The unknown question, however, is whether the instructional decision on either set of data (Daily or MW) was more educationally valid than the other. These data show only that instructional decisions were significantly different.

The third question in comparing Daily versus less-than-daily data graphs pertained to the type of program recommendations when changes in the instructional process were indicated. The lack of intrasubject reliability precluded this analysis.

Because data interpretation and data-based decision-making seem to be so integrally related it was quite surprising that the teachers could accurately and reliably determine how well the students were doing, but could not make accurate and reliable instructional decisions on those same data. It was particularly distressing because all of these subjects had received training in how to make data-based instructional decisions and apparently had been making decisions for some time in the classroom. It is possible that these problems with data-based decisions are unique to the particular program for educating teachers of the severely handicapped, and are indicative of not adequately teaching the rules for making accurate decisions. On the other hand, these problems may be due to the state of the art of decision-making and/or the willingness of teachers to make decisions based on data. If this latter explanation was true, problems in accuracy and reliability in the decision-making process might be expected across the country. In fact, the following observation made by Haring, Liberty and White (1980) at the University of Washington seems to lend some credence to the notion that it is quite difficult to get teachers to maintain instructional decision-making skills. They stated:

> "When classroom teachers began collecting data, they were taught to use visual analyses of charted or graphed performance data in order to make decisions about the effects of various instructional strategies in much the same way the researchers did. Unfortunately, individuals may interpret identical data differently. Following the development of simple analytical tools, such as a uniform method of summarizing change over time, the reliability of interpretation of visually inspected data improved. Even with such analytical tools, however, Liberty found that teachers trained in the use of such analytical tools did not use the data to make instructional changes in ineffective programs" (p. 160).

It is obvious that further research needs to be conducted on the questions of how many data should be collected, how data should be used, and the entire data analysis and decision-making process. Until such research studies are completed, professionals in teacher training programs need to be cautious not to be overzealous in the prescription of certain types of data systems.

REFERENCES

Browning, R. M. *Teaching the severely handicapped child*. Boston: Allyon & Bacon, Inc., 1980.

Conover, W. J. *Practical nonparametric statistics*. New York: John Wiley & Sons, 1971.

Edgar, E. Celebrate! How to bring out success in self-help skills. In P. W. Bailey (Ed.) *Ongoing data collection in the classroom*. Seattle: Western States Technical Assistance Resource, 1978.

Fredricks, H. D., Anderson, R., & Baldwin, V. Identifying competency indicators of teachers of the severely handicapped. *AAESPH Review*, 1979, *4*(1), 81-95.

Gentry, D. & Haring N. G. Essentials of performance measurement. In N. G. Haring & L. J. Brown (Eds.), *Teaching the severely handicapped*. (Volume I). New York: Grune & Stratton, 1976, 209-236.

Hanson, M. J. *Teaching your Down's Syndrome infant: A guide for parents*. Baltimore: University Park Press, 1978.

Haring, N. G., Liberty, K. A., & White, O. W. Rules for data-based strategy decisions in instructional programs: Current research and instructional implications. In W. Sailor, B. Wilcox, & L. Brown (Eds.) *Methods of instruction for severely handicapped students*. Baltimore: Paul H. Brookes, 1980, 159-192.

Mori, A. A. & Masters, L. F. *Teaching the severely mentally retarded: Adaptive skills training*. Germantown, MD: Aspen Press, 1980.

Sailor, W. & Haring, N. G. Some current directions in education of the severely/multiply handicapped. *AAESPH Review*, 1977, *2*(2), 67-87.

Wehman, P., Bates, P., & Renzaglia, A. Characteristics of an appropriate education for severely handicapped students. In P. Wehman & J. Hill (Eds.), *Instructional programming for severely handicapped youth: A community integration approach*. Richmond, VA: Virginia Commonwealth University, Department of Special Education, 1980.

Hey, Do We Really Have to Take Data?

Jennifer Holvoet, Ph.D.
Cathy O'Neil O.T.R.
University of Kansas

Lane Chazdin M.A.
Dian Carr
Janice Warner
Topeka Public Schools

1982

This paper was prepared as part of work pursuant to "Kansas Individualized Curriculum Sequencing Project" funded as a Handicapped Children's Model Project, U.S. Office of Education (G00-800-1721) from 9-1-80 to 9-1-83.

Abstract

Hey, Do We Really Have to Take Data

This paper used a counterbalanced ABA withdrawal design to look at several aspects of data collection on instructional programs that have been taught for more than one month. The data indicated: 1) that the teaching assistants could accurately teach such programs without using a data sheet to guide them through the task analysis; 2) on the majority of programs, the handicapped students learned as well under conditions where the teaching assistant collected no data as they did when data were collected; 3) two of the three teaching assistants made more accurate judgements of how well a student performed during a session when data were collected than when they were not; 4) all three of the teaching assistants made more accurate statements about how well a student's current performance compared to his previous performance when data were collected than when they were not; and 5) the teaching assistants also made better instructional decisions when data were collected. This study supports the literature that implies that the subjective judgements of teachers about student performance are more prone to error than those based on recorded data.

The use of daily data to measuring student progress in classrooms for the severely handicapped population has become the rule rather than the exception. Being skilled in collecting and analyzing data is generally thought to be one of the hallmarks of an effective classroom teacher.

Reasons for having teachers collect data are many and varied. For example, many authors (cf., Hanson, 1978; Liberty, 1976; Wehman, Bates, & Renzaglia, 1980) have stated that data should be taken to provide accountability that learning has taken place; to improve communication among parents, teachers, and administrators; and to allow the teacher to make instructional changes that will improve student's performance. Another common reason for collecting data is to provide an objective measurement of student behavior. Implicit in this reasoning is the belief that teachers are generally erroneous in their subjective judgments about student progress. This belief is present in a subtle form in many texts for special education teachers. For example, Van Etten, Arkell and Van Etten (1977) state:

> The more direct and continuous the data system employed, the more precise the teaching-learning process becomes... As the frequency and directness of measurement decreases, so does teacher knowledge of child learning behavior. (p. 343)

This quote implies that teachers are unable to tell how well a child is learning unless they use some sort of formal data system. This assumption is also present in a much more overt form in other texts. Consider for example, the following statements:

> There are many reasons why it is important to gather information or collect data on a child's progress in a training program. First of all, we all want the child to progress as rapidly as possible and therefore it is easy to think "Oh, yes, my child is doing fine." However, it is important for the child that we are very objective. . . (Hanson, 1978, p. 29).

> Direct measurement of skill acquisition, rather than inference or subjective judgement, is essential to the instructional system. Severely handicapped children have too much to learn in too short a time for teachers to waste time performing "ritualistic quasieducational" activities for which no varifiable student learning is demonstrable. Measurement tells you if the child is learning (acquiring the skill) and, equally important, tells you if he has learned (has "reached criterion"). (Sailor & Haring, 1980, 1981).

> Behavior is measured directly, not because therapists and parents deliberately present an inaccurate or unrealistic account of behavior, but because everyone has biases, or different ways of reporting what they see... For instance, often upon seeing a precise record of how often an individual emits a particular behavior, one staff member will exclaim, "I didn't realize he was doing it that often," while another will say, "Oh, I thought he did it much more often than that." To avoid misunderstandings, one directly observes and counts the client's behavior, and uses numbers to describe its frequency. (Favell, 1977, p.8)

Strangely enough, such statements about data collection do not appear to be data-based. There is little research verifying that teachers' subjective judgment about student progress are less accurate than their judgments based on data collected on a daily basis.

This study has investigated this issue for a very specific type of situation. Several teachers and paraprofessionals indicated to the authors that they felt that daily data collection was a waste of time when they had taught an instructional program to a student for more than one month. In other words, they felt that after one month of data collection on a given program, they could maintain the accuracy and precision needed to teach the skill and could make accurate subjective judgments about student progress. In the hopes of reducing their paperwork, they requested a data-based study that looked at several aspects of programs which had been taught for more than one month. The questions of interest were: 1) could they teach such a program accurately without using a data sheet to guide them through the various steps (i.e., did they skip steps and did the students continue to perform at the same or better levels); 2) were their judgments about student performance in a session as accurate under conditions where they took no data as they were under conditions where they did collect data; 3) were their judgments about how well a student's present performance compared to his previous performance accurate under conditions where no data was taken; and 4) could they make "good" programming decisions based on their subjective judgments?

METHOD

Subjects

Three teaching assistants from classrooms serving severely handicapped adolescents participated as subjects. Although several experienced teachers were interested in participating in this study, teaching assistants were chosen because their skills, both in data collection and their confidence in their subjective judgments, seemed to be more similar to those of a beginning teacher.

John was a 27 year-old male with two Bachelor of Arts degrees, one in music therapy and one in psychology. This was his first job working in a public school setting with severely handicapped children. He had been taught to collect data by the classroom teacher. Before beginning the study, he had been taking daily data on several instructional programs in the classroom for about four months.

Jill was a 24 year-old female who was in her Junior year at a local university. Her major fields of interest were speech pathology and education. When the study began, she had worked as a paraprofessional in a classroom for severely handicapped adolescents for slightly more than one year. Jill had been taught by a previous classroom teacher to record data and had been taking daily data on several instructional programs for the entire time she had worked in classrooms for the severely handicapped.

Denise was a 23 year-old female in her Junior year at a local university. Her major was health and recreation. At the time of the study, she had worked as a teaching assistant in a classroom for severely handicapped students for approximately one year. She had been taught to take data by her classroom teacher and members of a federally funded demonstration project associated with the SMH classroom in which she worked. She had been recording daily data on programs for several students throughout her career as a teaching assistant.

None of the subjects had been specifically taught to make instructional decisions from data, but had observed their teachers use data to make changes in instructional programs. Each of these subjects had asked to participate in the study because they felt that data collection, particularly on long-term programs, did not significantly enhance their teaching skills.

Setting

The subjects were observed as they taught specific instructional programs to a severely handicapped student on a 1:1 basis. The instruction took place in functional settings within a public school that served special populations. John was observed teaching dressing programs in the men's locker room. His student was seated on a bench for one of the programs and stood during the other. John usually stood or kneeled in front of the student while teaching. The observers were the only other persons present during John's teaching sessions.

Jill was observed in the home economics room where she and the student sat beside each other at a large table. Several other students and staff (as many as 15 persons on some occasions) were working in the room at the same time.

Denise was observed as she worked with a student in a classroom for severely handicapped students. The student was seated at a desk and Denise sat across from him. Five to ten other individuals (students and teachers) were working at two other tables in the same room.

Instructional Tasks/Students

Two criteria were used when selecting the instructional programs to be observed. First, the program had to have been taught by the teaching assistant, for at least ne month on a daily basis. Second, the student's performance on the instructional program needed to be variable in order to rigorously test whether the teaching assistants could accurately assess the student's performance.

John was observed teaching two self-help tasks to an 18 year-old severely handicapped male. This student was characterized as being highly distractible, aggressive, and noncompliant. His performance on all instructional tasks was highly variable, primarily due to his behavioral problems. The tasks selected for observation were: 1) putting on his shoes (this did not include tying the laces), and 2) putting on his pants. The Putting on Shoes program was task-analyzed into eight steps and Putting on Pants was task-analyzed into ten steps. All the steps of each program were taught daily. A three-step prompting procedure was used to teach the task. Verbal direction, modeling, and physical help were given sequentially until the student did a step correctly. These tasks were selected because John had been teaching these tasks to the student, on a daily basis, for about three months. Several program changes had been made during that time, but the student's performance remained highly variable and below criterion.

Jill was observed while she taught two daily-living tasks to a 17 year-old severely retarded female. Her student wore glasses to correct a visual defect and was characterized as difficult to manage due to stereotyped behavior, withdrawal from other people, and occasional severe temper tantrums. This student, however, was under good instructional control most of the time. The two tasks selected for observation were: 1) using a knife to cut bread, and 2) buttering bread. These tasks were selected because the student had not met criterion on them after four months of daily instruction, despite several changes in instructional method and materials. Each task was task-analyzed into four steps that were taught daily. Jill used the same three-step teaching strategy that John used.

Denise was observed teaching a prevocational sorting task to an 8 year-old, severely retarded male. This student was described as mildly self-injurious, highly dependent on physical prompts, and difficult to motivate. This student was given a four-compartment tray that had a different item in each of the three compartments across the top, and an object that matched one of the three items in the lower compartment. The student was to place the object from the lower compartment into the compartment with the matching object. A session consisted of five successive sorting trials. A graduated physical prompting system was used when the student did not independently and correctly do the task. Denise had been teaching this program for about three months when the study began. Two program changes had been tried during that time, but the student's performance had been highly variable. It was believed that the fluctuations in performance were related to motivational variables. One program modification (a change of materials) was made during the course of the study.

Procedures

Two different experimental conditions were used to compare the accuracy of each subjects' teaching and each subject's determination of student progress. In the DATA condition, the teaching assistant had a data sheet in front of him/her, and was asked to record the student's performance on each step or trial of the program. An observer, who sat near the teaching assistant and student, used an identical data sheet to record the student's performance. In the NO DATA condition, the teaching assistant was not allowed to use or look at the data sheet (or a list of the steps in the program) and did not record the student's performance. The observer recorded student performance on the tasks in the same manner used in the DATA condition.

The observer's data sheet, in both conditions, was used as the standard when determining the accuracy of the teaching process, the scores made by the handicapped students, and the accuracy of the teaching assistants' statements about the student's present and comparative performances.

To determine the accuracy of the teaching process in both conditios, the observer recorded how many steps or trials the teaching assistant skipped during the session. In the DATA condition, the teaching assistant had the steps of the program written on the data sheet. In the NO DATA condition, the paraprofessional had to remember the steps, or the number of trials that had elapsed. As a second measure of teaching accuracy, the handicapped student's performance scores were graphed under both conditions to determine whether data-recording by the subjects had any direct impact on the skill acquisition of the students.

Also measured was the ability of the teaching assistants to accurately judge how well the student had performed during a session and how this compared with previous performances. At the end of each session, the observer asked the subjects two questions:

1) "How well do you think your student did on this program today? Do you think he was at the bottom (0-20%), in the middle (21-79%), or at criterion (80-100%)?

2) "Do you think this student did better, worse, or about the same as he did the last time you taught him this program?" At the end of approximately every third session, the observer also asked:

3) "Do you think this program should be continued as it is, changed in some way, or should you go on to the next program or stop?"

The subjects were allowed to use their data sheets to answer these questions in the DATA condition and relied on subjective judgment in the NO DATA condition.

The accuracy of the paraprofessionals' judgements about the student's current performance (Question 1) and the student's comparative performance (Question 2) was determined by comparing the paraprofessionals' answers to these questions to the student performance data recorded by the observer. A subject's response judgment was judged to be accurate if the answer was in accord with the observer's data. For example, if the teaching assistant said the student was in the middle and the observer recorded a score of 30% correct, the subject was given credit for an accurate judgement. A subject was said to be underestimating if the answer indicated that the student did less well than indicated by the observer's data. For example, if the subject said a student was in the middle, but the observer recorded a 90% correct, the subject was scored as underestimating. Lastly, a subject was said to be overestimating if he judged the student to be doing better than indicated by the observer's data. For example, if the subject said the student was performing in the middle and the data indicated that the student had gotten a 19% correct, the subject was judged to be overestimating. No feedback was given to the teaching assistants about the accuracy of their answers until the study was concluded.

The accuracy of the teaching assistants' program decisions were determined by comparing the teaching assistants' answers to Question 3, with the decisions made by six persons who directed practica for teachers of severely multiply handicapped students at the University of Kansas. These "experts" were given graphic representations of the data available at the time the teaching assistant was asked to make a decision, and were asked to decide whether the program should be: (a) continued, (b) changed, or (c) discontinued or go on to the next step. For a teaching assistant's decision to be judged accurate, it had to agree with the decision of at least half (3) of the "experts."

Experimental Design

A single-subject ABA withdrawal design (Hersen & Barlow, 1976) was used in this study. The sequence of experimental conditions for Jill and Denise was: DATA, NO DATA, and DATA. John, on the other hand, had the reversed sequence: NO DATA, DATA, and NO DATA.

Reliability

Because the observer's data was used as a standard, high reliability scores between observers were considered very important. A second observer watched each program at least once (and usually 3 times) in every condition. This observer sat apart from the teaching assistant and the primary observer. Reliability data were collected on the primary observer's recording of student performance data, the number of steps skipped by the teaching assistant, and the teaching assistant's answers to the questions. Interobserver agreement for the primary observer's recording of student performance data was calculated using the exact-agreement method, where the data for each step or trial are compared (Repp, Deitz, Boles, Deitz, & Repp, 1976). An agreement was defined as a trial in which both observers agreed on the level of prompting used; disagreement as a trial in which the observers did not agree on the level of prompting. Interobserver agreement was determined by dividing the number of trials of agreement by the total number of trials in the session, and converting the result to percent.

The category method of calculating reliability (Repp, et al., 1976) was used for calculating agreement on the number of steps skipped. The observers looked at each step or trial taught and scored an agreement if both observers said the step was taught or if both observers said the step was not taught. A disagreement was scored if one observer scored a step as taught and the other did not. Reliability scores were determined by dividing the number of trials of agreement by the number of recorded trials and converting the result to percent.

The reliability observer's record of the answers made by the subjects to the questions was compared to the primary observer's record and the number of agreements was divided by the number of agreements plus disagreements. The quotient was then converted to percent.

10.

RESULTS

Reliability

Average reliability for the observer's recording of student performance data was 93%, 84%, and 92% for sessions conducted by Denise, Jill, and John, respectively. In the DATA condition, the average reliability for student performance was 90% during Denise's sessions, 84% for Jill's, and 100% for John's. The average reliability in the NO DATA condition was 100% for Denise's sessions, 81% for Jill's, and 90% for John's. The reliability for the number of steps/trials skipped by a subject was 100% in all cases. Interobserver agreement of 100% was also reached in all cases on the subject's answers to the three questions.

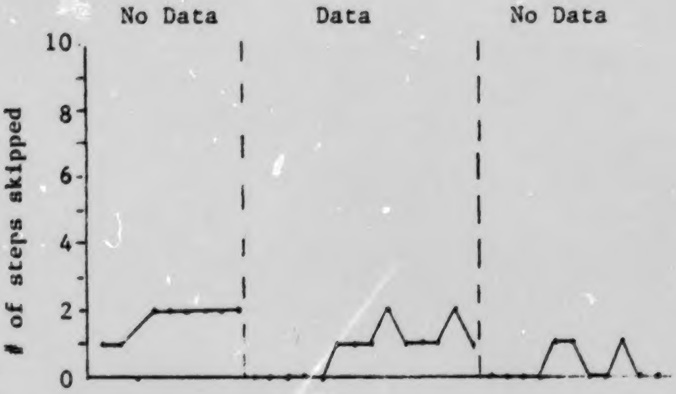
Number of Steps/Trials Skipped (Accuracy of Teaching)

In order to determine whether the recording of data had any effect on the accuracy with which a teacher followed a task analysis, the observer kept data on the number of steps or trials which the teaching assistant forgot to teach during each session. These data are presented in Figures 1 and 2. It should be noted that the task analyses were specified on the data sheet during the DATA condition, but the teachers did not have access to the task analyses during the sessions of the NO DATA condition.

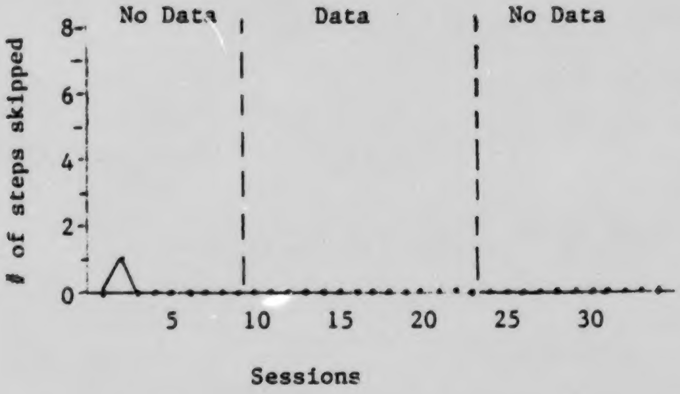
Insert Figures 1 & 2 about here

In the 10-step Pants On program taught by John, the mean number of steps skipped was 1.43, 1.09, and 0.3 in the NO DATA, DATA, and NO DATA conditions, respectively. In the eight-step Shoes On program taught by John, the mean number of steps skipped was 0.1, 0.0, and 0.0 in the NO DATA, DATA, and NO DATA conditions, respectively.

Figure 1. Number of steps/trials skipped by John during teaching sessions

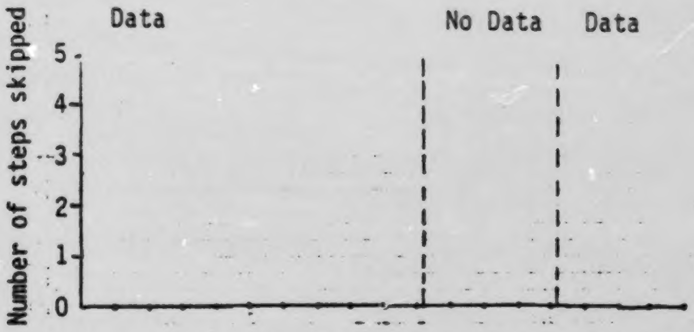


Teacher: John
Student: Bill
Program: Pants On

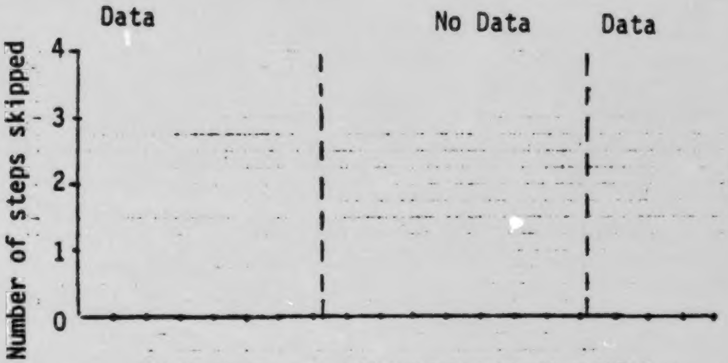


Teacher: John
Student: Bill
Program: Shoes On

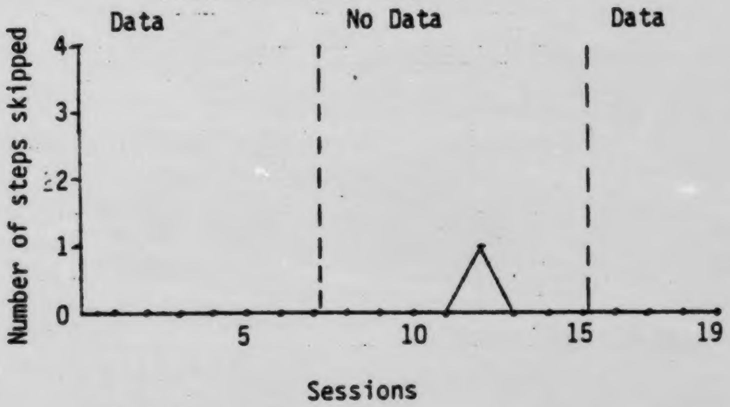
Figure 2. Number of steps/trials skipped by Denise and John during teaching sessions



Teacher: Denise
Student: Dick
Program: Matching Shapes



Teacher: Jill
Student: Marge
Program: Butter Bread



Teacher: Jill
Student: Marge
Program: Cut with knife

Denise (Figure 2) consistently did all five trials of the matching program, regardless of experimental condition.

Likewise, Jill (Figure 2) skipped no steps of the 4-part sequence of the Buttering Bread program in either condition. She did, however, leave out one step of the 4-step program for Use of Knife during one session under the NO DATA condition (for a mean of 0.1 skipped steps). The mean skipped steps during the DATA condition on this program was 0.0.

Effect of Data Recording on Student Performance

The learning curves generated by the handicapped students are presented in Figures 3 for John and 4 for Denise and Jill. John's student, Bill, shows a decreasing trend in the first NO DATA condition of the Shoes On program. He then shows an accelerating trend in the DATA condition, and a fairly stable performance above criterion level during the second NO DATA condition. The means in this program were 63.0% in the first NO DATA condition, 79.1% in the DATA condition, and 81.3% in the second NO DATA condition. In the Pants On program, Bill showed a quickly decelerating curve in the first NO DATA condition, and variable performance in the DATA and second NO DATA condition. The mean performances were higher in the DATA condition (68.0%) than those obtained in the two NO DATA conditions (60.0% & 58.2%).

Insert Figures 3 & 4 about here

Denise's student (Figure 4) showed accelerating trends under both DATA conditions, but a flat trend under the NO DATA condition. The mean were 47.5% in the first DATA condition, 60.0% in the NO DATA condition, and 80.0% in the second DATA condition.

Figure 3. Percent of steps taught by John completed independently by student

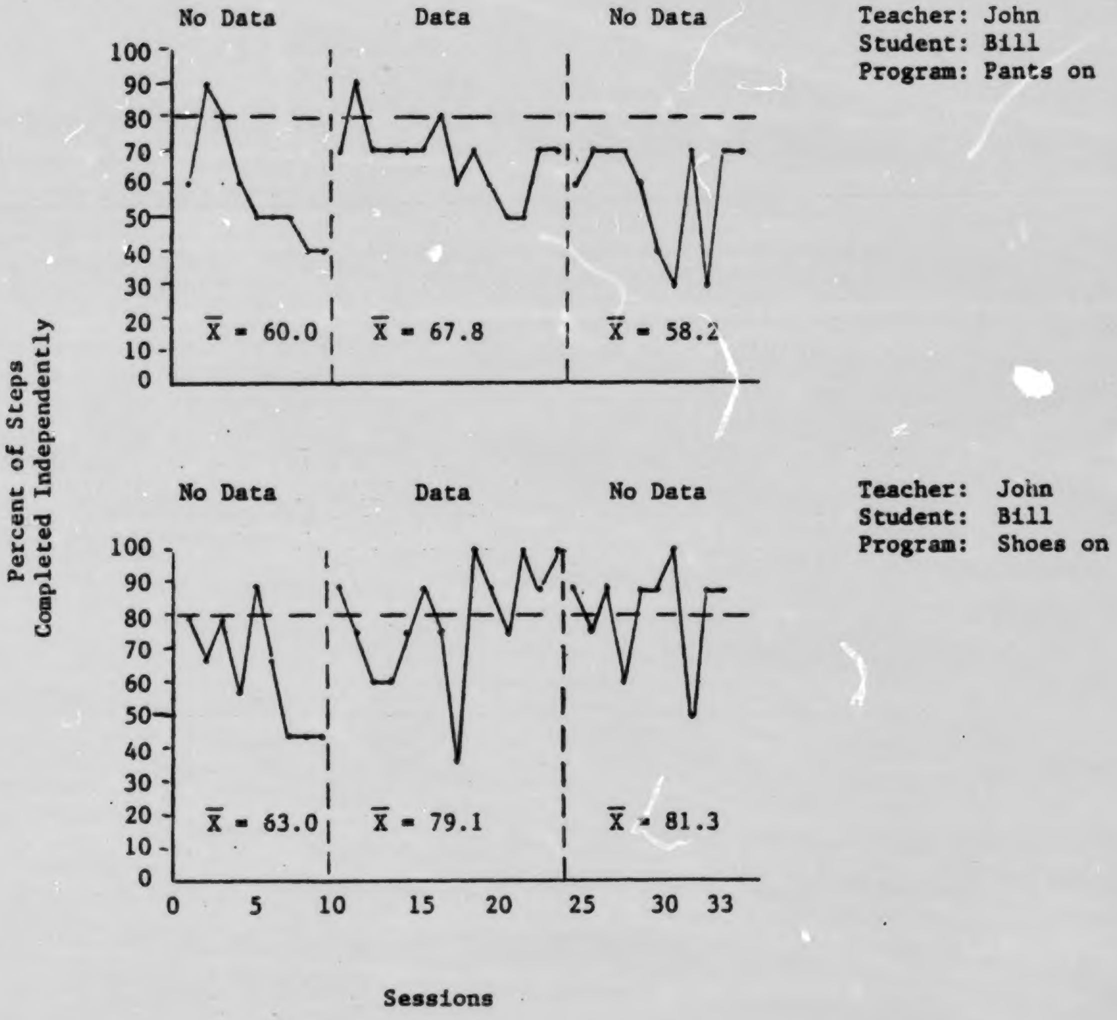
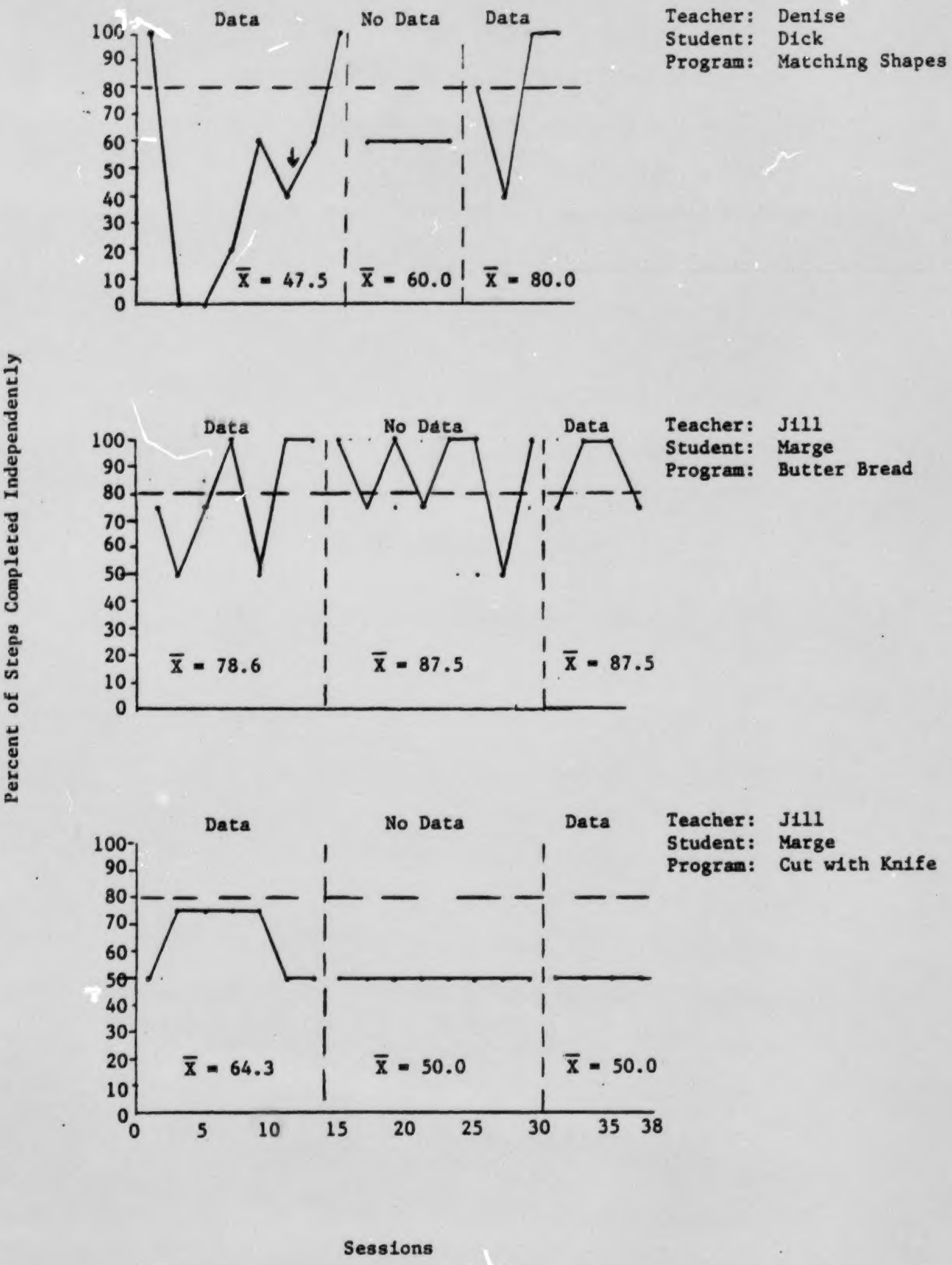

No Data
Data
No Data
Teacher: John
Student: Bill
Program: Pants on
100
90
80
70
60
50
40
30
20
10
0
X̄ = 60.0
X̄ = 67.8
X̄ = 58.2
Percent of Steps
Completed Independently
No Data
Data
No Data
Teacher: John
Student: Bill
Program: Shoes on
X̄ = 63.0
X̄ = 79.1
X̄ = 81.3
0
5
10
15
20
25
30
33
Sessions

Figure 4. Percent of steps taught by Denise and Jill completed independently by students


Data
No Data
Data
Teacher: Denise
Student: Dick
Program: Matching Shapes
X̄ = 47.5
X̄ = 60.0
X̄ = 80.0
Data
No Data
Data
Teacher: Jill
Student: Marge
Program: Butter Bread
X̄ = 78.6
X̄ = 87.5
X̄ = 87.5
Data
No Data
Data
Teacher: Jill
Student: Marge
Program: Cut with Knife
X̄ = 64.3
X̄ = 50.0
X̄ = 50.0
0
10
20
30
40
50
60
70
80
90
100
0
5
10
15
20
25
30
35
38
Percent of Steps Completed Independently
Sessions

Jill's student (Figure 4) showed high, but variable performance, on the Buttering Bread program regardless of experimental condition. The means were 78.6%, 87.5%, and 87.5% for the DATA, NO DATA, and DATA condition, respectively. The student showed stable performance on the Cut with Knife program, regardless of experimental condition. The means were 64.3%, 50.0% and 50.0% for the DATA, NO DATA, and second DATA conditions, respectively.

Overall, there were lower performance scores in the NO DATA condition(s) for the Pants On program and the Matching Shapes program.

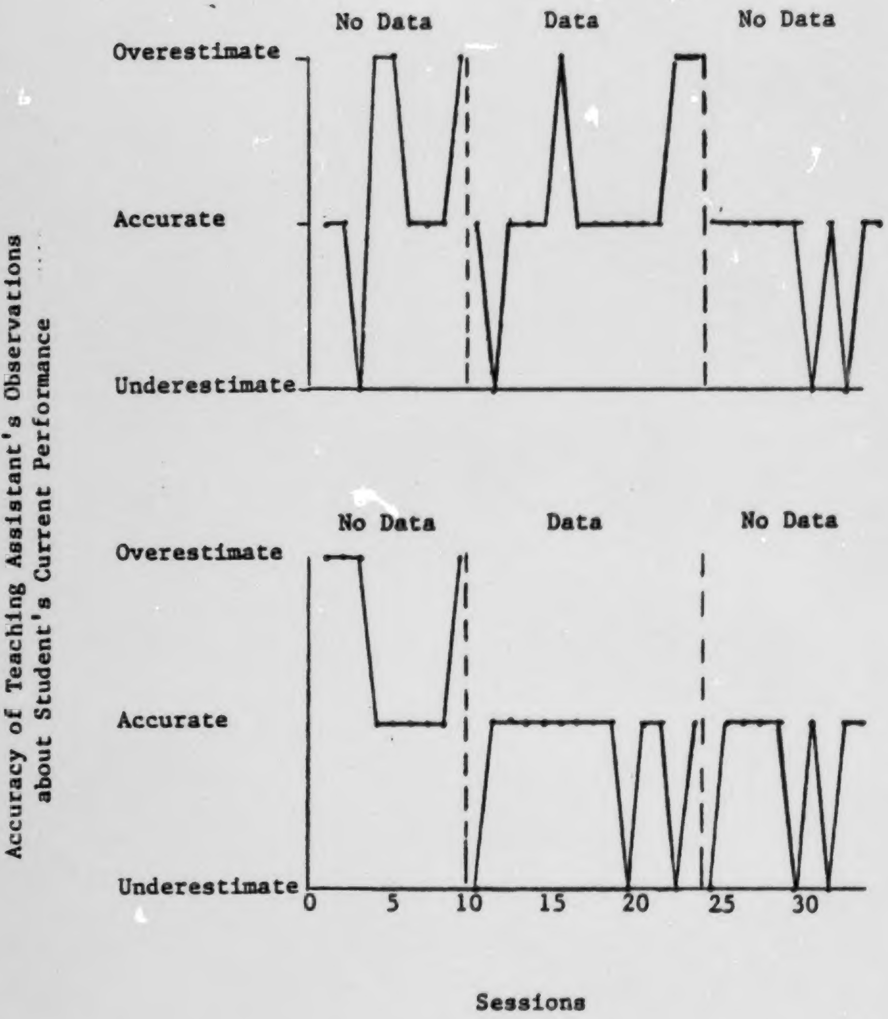
Accuracy of Statements about Current Student Performance (Question 1)

Figures 5 and 6 illustrate the accuracy of each teaching assistant's determination of how well a student did in a given session. In the Pants On program, John made 14 accurate statements (i.e., his statements agreed with the observer's data) in 20 sessions (70% accuracy) during the NO DATA condition. He made 10 accurate statements in 14 sessions (71% accuracy) during the DATA condition. When he taught the Shoes On program, he was accurate on 12 out of 19 sessions (63%) in the NO DATA condition and 11 out of 14 sessions (79%) in the DATA condition. Interestingly enough, in the first NO DATA condition on both programs, the majority of his errors were in the direction of overestimation, but in the second NO DATA condition, his errors were always in the direction of underestimation.

Insert Figures 5 & 6 about here

Denise (Figure 6) was accurate on 3 out of 4 NO DATA sessions (75% accuracy) and was accurate on 100% of the DATA sessions. She overestimated the student's performance when the error was made.

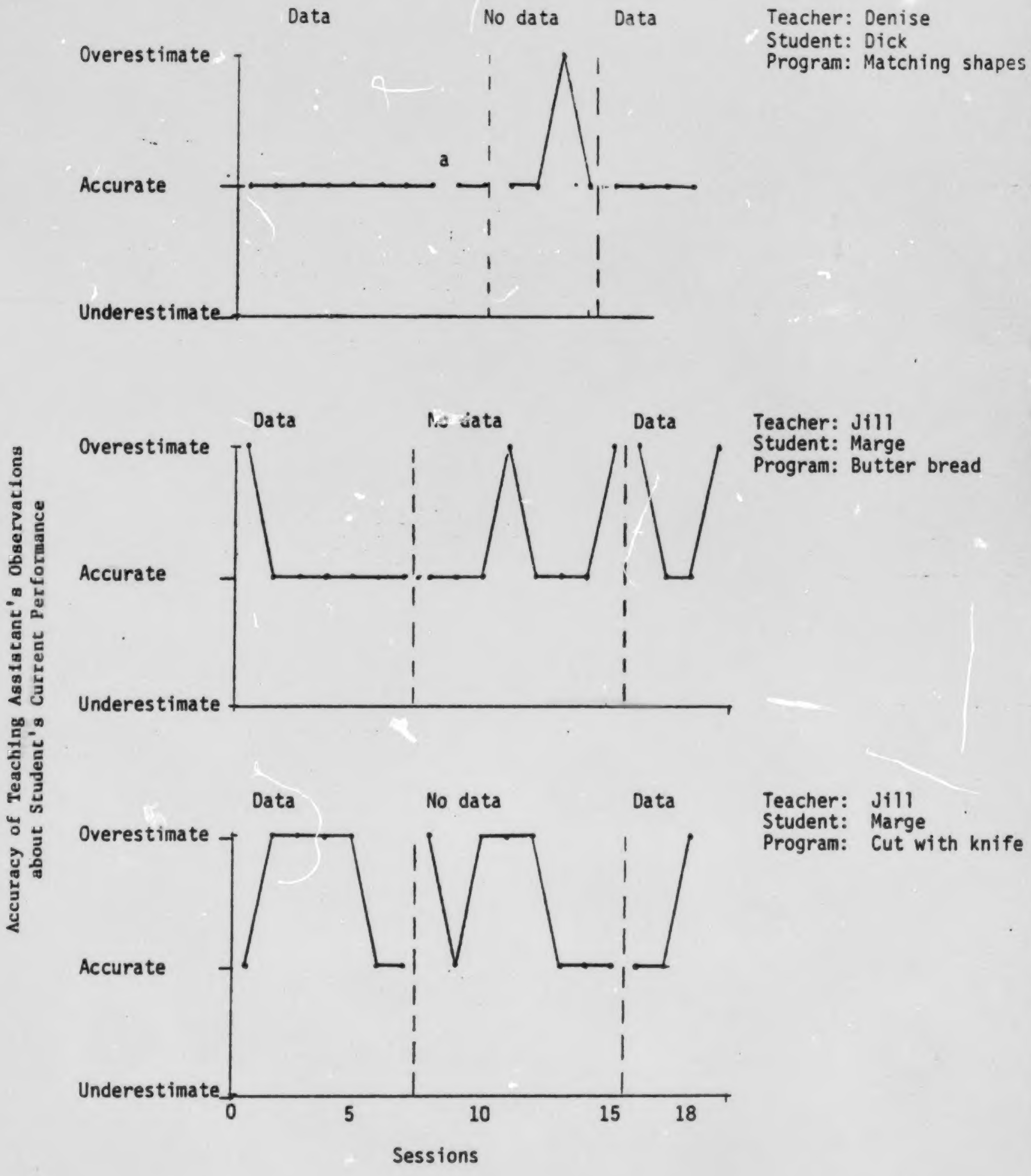
Figure 5. Accuracy of John's statement about current student performance (Question 1)



Teacher: John
Student: Bill
Program: Pants on

Teacher: John
Student: Bill
Program: Shoes on

Figure 6. Accuracy of Denise's and Jill's statements about current student performance (Question 1)



a= change of materials used in program.

In the Buttering Bread program (Figure 7), Jill made accurate statements on 6 out of 8 sessions (75%) during the NO DATA condition, and on 8 out of 11 sessions (73% accuracy) during the DATA condition. When she taught the Cut with Knife program, the NO DATA condition resulted in accurate statements on 4 out of 8 sessions (50%); similarly the DATA condition resulted in 5 accurate statements out of 10 sessions (50% accuracy). In both programs, Jill showed a pattern of overestimating the student's performance.

Overall, when asked to estimate how well the student did during a session, John and Denise were somewhat less accurate during the NO DATA condition. Jill, on the other hand, showed equivalent accuracy under the two conditions for one program and was more accurate in the NO DATA condition for the other program.

Accuracy of Statements about Student's Comparative Performance (Question 2)

Figures 7 and 8 illustrate the accuracy of the subject's responses to Question 2 where they were asked to compare the student's performance on a given session to the performance during the previous session. John had only 18 recorded days in the NO DATA condition on the Pants On program and 17 recorded days of NO DATA in the Shoes On program because it was impossible to assess the accuracy of the comparative judgement on the first day of the study, and because the observer forgot to ask Question 2 on Session 9. He had only 10 days in the DATA condition for the Shoes On program because the observer forgot to ask Question 2 on the first four days of this condition. John's statements about the

----

Insert Figures 7 & 8 about here

Figure 7. Accuracy of statements about comparative student performance (Question 2)

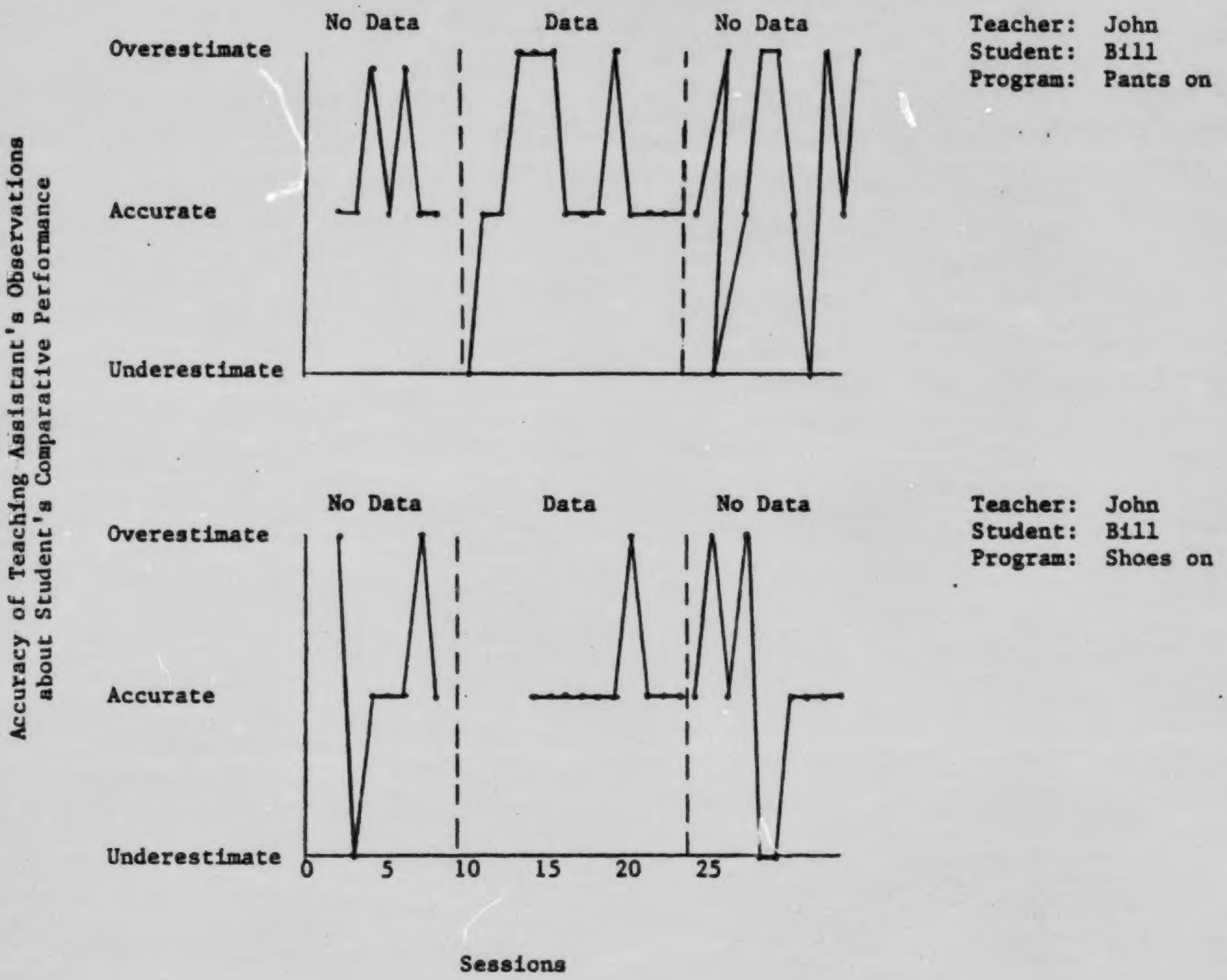
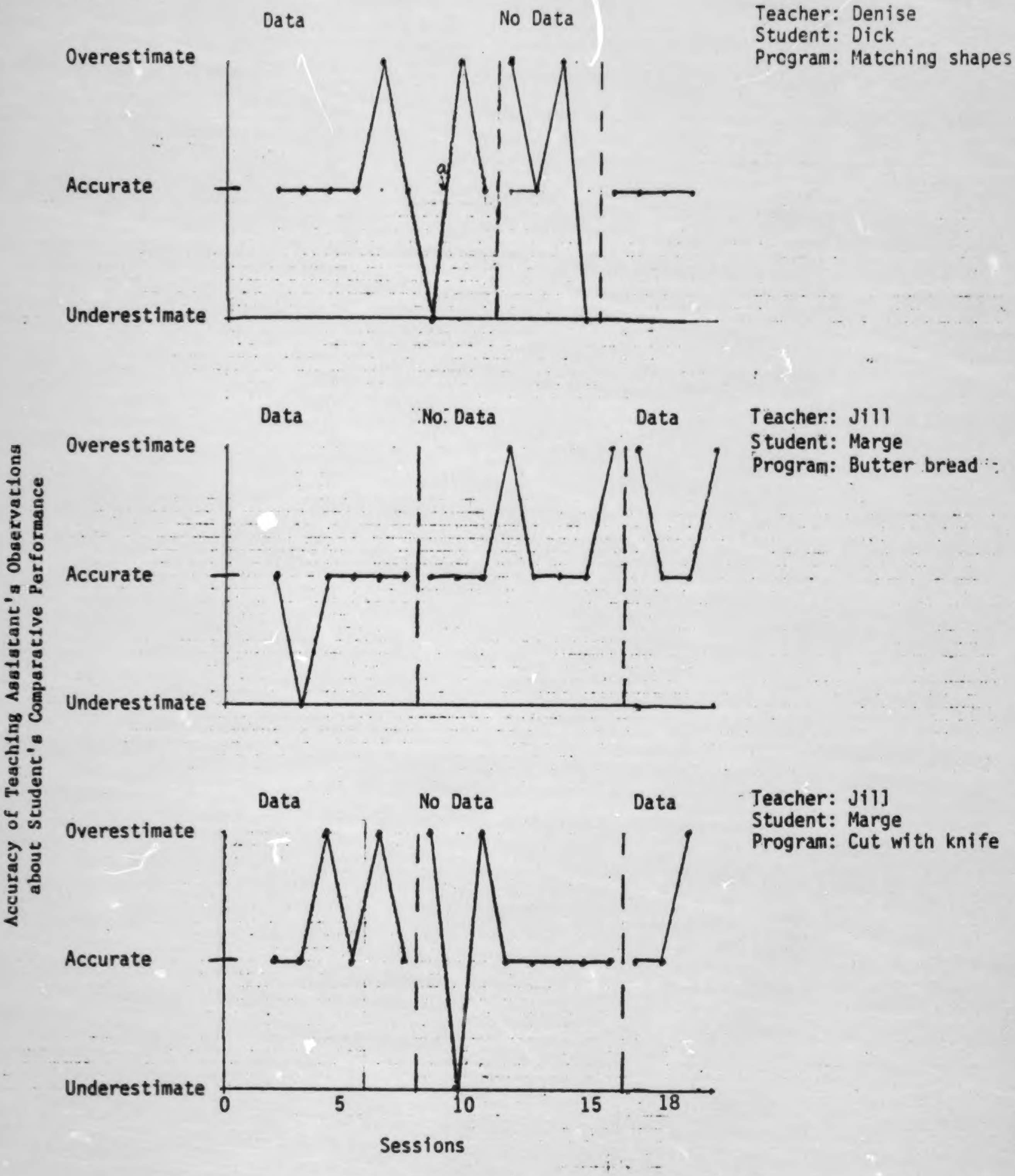

Accuracy of Teaching Assistant's Observations about Student's Comparative Performance
No Data
Data
No Data
Overestimate
Accurate
Underestimate
Teacher: John
Student: Bill
Program: Pants on
No Data
Data
No Data
Overestimate
Accurate
Underestimate
0
5
10
15
20
25
Teacher: John
Student: Bill
Program: Shoes on
Sessions

Figure 8. Accuracy of statements about comparative student performance (Question 2)



a= change of materials used in program.

student's comparative performance on the Pants On program were accurate on 9 out of 18 sessions (50%) when the NO DATA condition was in effect. He was accurate on 9 out of 14 sessions (64%) during the DATA condition. His comparative statements about the Shoes On program were accurate on 10 out of 17 (59%) NO DATA sessions, and on 9 out of 10 (90%) DATA sessions.

Denise (Figure 8) made accurate comparative statements on 1 out of 4 (25%) of the NO DATA sessions and on 10 of the 13 (77%) DATA sessions.

On the Buttering Bread program (Figure 8), Jill was accurate on her comparative statements on 6 out of 8 sessions (75%) during the NO DATA sessions and on 7 of the 10 (70%) DATA sessions. In the Cut with Knife programs, she was accurate on 5 of the 8 (63%) NO DATA sessions, and on 6 of the 9 (66%) DATA sessions.

In summary, with the exception of the Buttering Bread program, the teaching assistants made comparative judgments about the student's performance more accurately under the DATA condition.

Accuracy of Instructional Decisions (Question 3)

The accuracy of the teaching assistant's decisions about whether the instructional program should be continued, changed or discontinued/go to next step is illustrated in Table 1. The first three columns

Insert Table 1 about here

represent the subject's name and instructional program, the session number on which the decision was made, and the experimental condition in effect during that session. Shaded boxes indicate the decisions made by the teaching assistant. The numbers in each of the boxes indicate how many of the "experts" (N=6) decided to continue, change, or discontinue

TABLE 1

Comparison of Data Decisions Made
By Six "Experts" and Paraprofessionals

Subject/Instructional Program	Session #	Experimental Condition	Instructional Decisions[a]		
			Continue	Change	Discontinue or Go to Next Step
John - Pants On	3	No Data	6 *	0	0
	5	No Data	0	6	0
	8	No Data	0	6 *	0
	10	Data	2	4	0
	13	Data	3 *	3	0
	17	Data	0	6 *	0
	20	Data	0	6	0
	23	Data	1	5	0
	26	No Data	0	6	0
	28	No Data	0	6	0
	31	No Data	0	6	0
	34	No Data	0	6	0
John - Shoes On	3	No Data	5 *	1	0
	5	No Data	5	1	0
	8	No Data	0	6	0
	10	Data	3 *	3	0
	13	Data	0	6	0
	17	Data	1	5 *	0
	20	Data	1	1	4
	23	Data	0	0	6
	26	No Data	0	0	6
	28	No Data	0	0	6
	31	No Data	0	0	6
	34	No Data	0	0	6
Denise - Matching Shapes	2	Data	3 *	3	0
	5	Data	6 *	0	0
	8	Data	6 *	0	0
	11	No Data	0	6	0
	14	Data	0	6	0
	16	Data	6 *	0	0

			Continue	Change	Discontinue or Go to Next Step
Jill - Buttering Bread	2	Data	5 *	1	0
	5	Data	6 *	0	0
	9	No Data	1	0	5
	12	No Data	0	0	6
	15	No Data	0	0	6 *
	18	Data	0	0	6 *
Jill - Cut with Knife	2	Data	6 *	0	0
	5	Data	3 *	3	0
	9	No Data	0	6	0
	12	No Data	0	6 *	0
	15	No Data	0	6 *	0
	18	Data	0	6 *	0

Note: Shaded areas represent decision made by paraprofessional and numbers represent the number of experts choosing each category.

a : Stars indicate an agreement between the paraprofessional's decision and the decision of at least 3 of the experts.

the program based on a graphic representation of the data up to that point. For example, in Session 2 for Denise's student's data, three "experts" thought the program should be continued, three thought the program should be changed, and none thought it should be discontinued. Denise, on this same session, thought the program should be continued.

As can be seen in Table 1, Denise agreed with at least half of the experts on 4 of her 5 decision points (80%) made under the DATA conditions, but did not agree with the experts on the one decision she made under the NO DATA condition. Jill agreed with at least half of the experts 100% of the time (3 out of 3) under the DATA conditions for both programs. She agreed with at least half of the experts only 33% (1 out of 3) and 67% (2 out of 3) times under the NO DATA condition for the Buttering Bread and Using Knife programs, respectively. John, in the Pants On program, had 29% agreement (2 out of 7) in the NO DATA condition and 20% agreement (1 out of 5) in the DATA condition. The decisions John made on the Shoe On program were made by at least half the experts on 1 of the 7 decision points (14%) in the NO DATA conditions, and on 2 out of the 5 decision points (40%) of the DATA conditions.

Overall, on four of the five programs, the teaching assistants made better decisions (i.e., they agreed with the decisions made by at least half the experts) under the DATA condition.

DISCUSSION

The first issue of concern in this study was whether the teaching assistants could accurately teach a long-term program without skipping any of the steps/trials when they did not record data or have access to the task analysis during the session. The study indicated that the accuracy of following a task-analysis was surprisingly good under conditions of NO DATA. Both Denise and Jill were able to teach a short program without error. On Jill's second program, she only skipped a step once in the eight days she taught under the NO DATA condition. John also showed a great deal of similarity between the accuracy of his teaching in the two conditions. In the Shoes On program, he made only one error in 18 NO DATA sessions. This seems to be quite accurate considering there were 8-steps in this program. In the Pants On program, John consistently left out one step in both conditions. He verbalized that he felt this step was superfluous. He also left out a second step intermittently under the first NO DATA and the DATA conditions. Thus, it would appear that although the teaching assistants occasionally left out a step during the NO DATA condition, they maintained a level of accuracy very close to that obtained in the DATA condition. This occurred even though they were not allowed access to the task analysis under the NO DATA condition. It is possible, and even probable, that this degree of accuracy in the NO DATA condition would not have been obtained on a new instructional program. It is probably not too surprising that the teaching assistants were able to remember the task-analyses of programs that they had been teaching in a data-based structure for several months.

A second measure of teaching accuracy might be inferred from the student performance data. If the teaching were equally accurate under both conditions, one might expect to see an accelerating learning curve across conditions, or equal performance across conditions. If on the other hand, teaching accuracy (or quality) was less effective in the NO DATA condition, one might expect to see poorer student performance in the NO DATA sessions. Looking at the mean scores, it can be seen that three of the programs, Shoe On, Matching Shapes, and Buttering Bread, showed a pattern of learning across sessions regardless of condition. Thus, it might be inferred that teaching accuracy was similar across conditions for these programs. The Cut with Knife programs showed a decelerating trend beginning near the end of the first DATA condition and remaining stable across the other two conditions. It is difficult to infer anything about the teaching accuracy from these program data. The other program, Pants On showed clearly lower scores under the NO DATA condition. This would seem to imply less accuracy on some dimension of teaching in the NO DATA condition of this program.

The second area of interest was whether the statements made by the teaching assistants about how well a student did in a session (Student's current performance) were as accurate under the NO DATA condition as they were under the DATA condition. This investigation showed that John and Denise made accurate statements more often under the DATA condition than they did under the NO DATA condition. Jill, on the other hand, was equally accurate (or inaccurate) in the two conditions of the Cut with Knife program, and was slightly more accurate on the NO DATA condition for the Buttering Bread program. One of the most interesting findings was that both John and Jill had surprisingly low accuracy in the DATA condition. Since they could refer to their data sheet and see how well the student did, it would seem that they should have been much closer to 100% accuracy. Three things could account for these low scores. First, it is possible that the primary observer was not accurate. However, reliability scores comparing the primary observer's and the reliability observer's student performance data were uniformly above 80%. The second possibility is that the teaching assistants just miscalculated, since the questions were phrased in terms of a percent, but the subjects did not have access to a calculator. It would seem that this is a likely explanation for at least some of the error, since both Jill and John would sometimes say things like "Well, s/he got three out of four so I guess s/he's at criterion." Another possible source of error in the DATA condition is that the teaching assistants did not use strictly objective judgments of the students' performance. This also is a likely explanation. John would sometimes score the student as performing independently on the steps that were not even taught (i.e., the skipped steps). Similarly, Jill would often give the student credit for cutting the bread independently when the bread did not come apart after the student had made several efforts to cut it. Although there was evidence that taking data improved accuracy, the similarity of scores in the two conditions was surprising. This, to some extent, upholds the belief of the teaching assistants that they were just as accurate in determining how well a student was doing regardless of whether they took data or not. On the other hand, it also lends some credence to the professionals' stand that recording data does improve a teacher's ability to know exactly how well a student is doing. It is also important to note that it is possible that these subjects might have been far less accurate in the NO DATA condition if they had not had a history of data recording.

The third area of concern in the study was whether the statements paraprofessionals made about students' comparative performances were as accurate under NO DATA conditions. In four of the five programs, the comparative statements were more accurate under the DATA condition. Only Jill's Buttering Bread program was more accurately assessed under the NO DATA condition. It should be noted that the demand characteristics (Rosenthal, 1966; 1967) of this study, which made the subjects focus on how well the student did each day and also try to determine comparative performance on a daily basis, might have inflated the NO DATA scores over the accuracy that would normally be obtained in the field where such demand characteristics would not be present. Therefore, it appears that these data support statements made by several professionals that repeated measurement gives a more accurate picture of a student's overall functioning than does subjective judgement.

The fourth issue addressed in this study was whether recording data improved the instructional decision-making process. Both Jill and Denise demonstrated the ability to make accurate data decisions (i.e., agree with at least half of the "experts") in the DATA condition even though they had never had specific training in instructional decision-making. Under the NO DATA conditions, however, their accuracy dropped considerably. John demonstrated considerable problems with the decision-making process under both experimental conditions. Overall, on four of the five programs, better decisions were made under the DATA conditions. These findings are in accord with the professional consensus (Haring, 1978; Perske & Smith, 1977; Van Etten, Arkell, & Van Etten, 1980; Whelan, 1972) that repeated measurements help teachers make good instructional decisions.

In summary, this study seems to support the literature that implies that the subjective judgments of teachers about student performance are more prone to error than those based on recorded data (cf., Favell, 1977; Hanson, 1978; Sailor & Haring, 1980; Wehman, Bates, & Renzaglia, 1980). Although the subjects in this study were able to teach a long-term program fairly accurately without a data sheet and their accuracy in judging a student's current performance was similar under both conditions, it was apparent that comparative judgements and decision-making were adversely affected when data were not recorded.

Whether these findings would be representative of teachers of the severely handicapped who have been extensively trained in data collection and analysis is not clear. It is hypothesized that the differences between the DATA and NO DATA conditions might be even larger with more highly trained subjects because they might show higher accuracy in the DATA condition than did the subjects in the current study. It is also important to recall that these data are based on long-term instructional programs. Therefore, the subjects in this study had a long history of data collection and teaching on the instructional programs used in the study. The effect of this history on the results of this study are unknown, though it is hypothesized that such a history should have improved their ability to teach and judge student performance accurately. Further inquiry needs to be made to determine whether similar results are found when the subjects do not have a data-recording history, enjoy the data-recording process and/or when the programs are new to the teacher and student.

References

Favell, J. E. The power of positive reinforcement: A handbook of behavior modification. Springfield, IL: Charles C Thomas, 1977.

Hanson, M. J. Teaching your Down's syndrome infant: A guide for parents. Baltimore, MD: University Park Press, 1978.

Liberty, K. Data dilemma: Response and measurement units for teachers of the severely handicapped. AAESPH Review, 1976, 1(5), 13-31.

Sailor, W. & Haring, N. G. Some current directions in education of the severely/multiply handicapped. AAESPH Review, 1977, 2(2), 67-87.

Van Etten, G., Arbell, C., & Van Etten, C. The severely and profoundly handicapped: Programs, methods, and materials. St. Louis, MO: C. V. Mosby Co., 1980.

Wehman, P., Bates, P., & Renzaglia, A. Characteristics of an appropriate education for severely handicapped students. In P. Wehman & J. Hill (Eds.) Instructional programming for severely handicapped youth: A community integration approach. Richmond, VA: Virginia Commonwealth University, Department of Education, 1980.

The Effects of Two Different Data Sheet Formats on Prompting Accuracy

Jennifer F. Holvoet, Ph.D.
University of Kansas
Lawrence, KS

Nancy Schussler, M.Ed.
University of Kansas
Lawrence, KS

Jerry Rea, B.A.
Kansas State School for the Deaf
Olathe, KS

Laurie O'Shea
Washburn University
Topeka, KS

Liz Vogt, M.S.Ed.
U.S.D. 501
Topeka, KS

1983

This paper was prepared as part of the work pursuant to the "Kansas Individualized Curriculum Sequencing Project" funded as a Handicapped Childrens' Model Project, U.S. Office of Education (G00-800-1721) from 9-1-80 to 9-1-83.

For at least a decade, measurement of student performance has been recognized as an important component of the instructional process in classrooms for students with severely handicapping conditions (Favell, 1977; Hanson, 1978; Liberty, 1976; Mori & Masters, 1980; Sailor & Guess, 1983; Snell & Smith, 1978; Sternberg & Adams, 1982; Wehman, Bates, & Renzaglia, 1980; White & Liberty, 1976). Proponents of classroom data collection and analysis systems have cited several reasons why data systems should be implemented. These include: 1) assessing the extent of the problem; 2) providing proof that learning is taking place; 3) providing a monitoring system that helps a teacher determine when an instructional program is ineffective and needs to be modified; and 4) improving communication among parents, teachers, and administrators. In spite of these good reasons, it is widely recognized that many teachers of the severely handicapped population do not use data systems or do so grudgingly (Haring, Liberty, & White, 1980; Sailor & Guess, 1983). In fact, a recent survey by the Association for the Severely Handicapped ("Say It Ain't So," 1981) showed that less than half of the teachers who responded evaluated progress on all of their students' programs and a similar proportion reported not using any rules to make instructional decisions.

Professionals have speculated about the factors causing this discrepancy between theory and practice. Sailor and Guess (1983) felt that "the primary obstacle to utilizing a data management system is the extent to which it competes for time with the instructional process" (p. 151), and that the system must be simplified before teachers will use it. This observation is probably accurate, but before changing the system it is important to determine if there are other variables which make it fail. One such variable appears to be that while a data collection system requires that the teacher use part of each instructional session to collect data, the benefits of the system are salient on far fewer occasions. For example, our observations have been that the majority of teachers do not communicate with another professional or a parent about a student's performance on a specific program more than once a month, except in the most unusual circumstances. Thus, even if the teacher were reinforced for using data on such occasions (and this is not always the case), there is a long delay between the required response (taking data) and the reinforcement. Similarly, at least 10 days of data are needed for good instructional decisions; thus decisions are usually made at two-week or longer intervals. Furthermore, making such decisions may mean that the teacher will need to spend extra time changing the instructional program, changing the data sheets, and communicating these changes to other staff. This further decreases the time available for direct instruction and, thus, may be perceived as punishing, rather than reinforcing. Laws of reinforcement specify that reinforcers work best if they occur frequently, if they occur immediately after the response has been emitted, and if there is not a lot of discrepancy between the amount of work required and the amount of reinforcement attained (Craighead, Kazdin, & Mahoney, 1976; Miller, 1975; Thompson & Grabowski, 1972). Since the postulated benefits of a data collection system violate all of these laws, we probably should not be surprised that teachers do not use the system more often.

On the other hand, a data collection system may have some immediate benefits (which may be reinforcing) which professionals have not identified. If it could be shown that the process of collecting data improved a teacher's instructional skills during a teaching session, teachers might be more persuaded of its value. This study is designed to determine if teaching accuracy, as measured by the correct use of a least-prompts correction strategy, is affected by the format of the data sheet used while teaching.

METHOD

Subjects/Observers

Three individuals participated as subjects in this study. These subjects had differed in their degree of experience with teaching the severely handicapped population, experience with data collection, and in their educational background. They were felt, by the authors, to be representative of three of the many different groups of staff that often work with severely handicapped students in a school setting: 1) an experienced teaching assistant; 2) an inexperienced teaching assistant; and 3) a highly experienced teacher.

Denise was a 24 year old female in her senior year at a local university. Her major was health and recreation. At the time of the study, she had worked as a teaching assistant in a classroom for severely handicapped students for approximately two years. She had been taught to conduct instructional sessions and to collect data by the classroom teacher and members of a federally funded demonstration project associated with the classroom in which she worked. She had been recording daily data on instructional programs since she began working as a teaching assistant.

Sam was a 30 year old high school graduate. At the time of the study, he had worked as a teaching assistant in a classroom for severely handicapped students for approximately six months. He had been taught to conduct instructional sessions and to collect data by the classroom teacher and had been taking daily data on several instructional programs for approximately four months.

Lisa was a 30 year old female working on her Master of Education degree at a nearby university. Her major was Special Education for the Severely Handicapped. She had worked as a primary teacher of nonhandicapped children for one year and had nine years of experience with children who were diagnosed as mildly to severely mentally retarded. She had been taught to conduct instructional sessions and collect data through her college classes and a federally funded language research project at a state residential facility for retarded children. She had been collecting data in research and classroom settings for approximately seven years.

All the subjects had used data recording formats that were similar, but not identical, to the formats of the data sheets used in the study.

Before beginning the study, the subjects were told that the investigators were trying to determine whether the type of data sheet an instructor used had any effect (good or bad) on teaching. They were also told that they would use two types of data sheets during the study. Once the study was underway, only the observers were allowed to see the information being collected, and they were told not to answer any questions about the study while it was ongoing. The instructors (subjects) were, however, allowed to visit with other staff about the instructional programs and to make changes in the programs if needed. When the study was completed with all subjects, each instructor was shown the data relating to his/her own accuracy in using the least-prompts strategy.

Two teaching assistants from a federally funded demonstration project served as observers. Both of these observers had worked in classrooms for severely handicapped individuals for at least three years and had used the least-prompts strategy extensively. One observer was a male and the other was female.

Setting

The instructors were observed as they taught specific instructional programs to one or two severely handicapped student(s). The instruction took place within a classroom in a public school that served special populations. The student(s) were seated at either a table or a student desk and the instructor sat across from, or next to, the student. Five to ten other individuals (students and staff) were working at other tables in the classroom while the study was being conducted.

Tasks/Materials

Each of the instructors taught two programs during the study. Denise was observed teaching a group of two severely handicapped students. She taught each of the students the same skill (Receptive Object Identification) and alternated trials between the two students. A session was comprised of 10 trials of both students' programs, making a total of 20 trials.

Sam taught one student two different tasks (Matching Objects and one-part Kit Assembly) using an alternating format where first he taught a trial of Matching and then he taught a trial of Kit Assembly. A session was comprised of 12 trials of each program making a total of 24 trials.

Lisa also taught one student two different tasks (One-step Direction Following and Placing "One") using the alternating format described above. Her sessions were comprised of 10 trials of each program, making a total of 20 trials.

Teacher Strategy

All the observed programs were taught using the least-prompts teaching strategy outlined by Lent (1975), no matter which data sheet format was used. The least-prompts strategy is a systematic hierarchy of prompts that gradually increases the amount of assistance given by the instructor whenever the student responds incorrectly or does not respond. Use of this strategy specifies four levels of prompting that are presented in a designated order until the student makes an acceptable response. These levels consist of: 1) I = an initial cue is given, such as "Find the same"; 2) V = a verbal prompt is given, such as "Find another cup"; 3) D = a demonstration of the desired response is given, such as the instructor pointing to the cup(s); and 4) P = the student is given physical assistance to complete the task, such as the instructor moving the student's hand toward the correct object.

When the strategy is implemented correctly, the instructor should progress through the prompting levels in the exact order specified until the student makes a correct response, and should give no more than one prompt at each level. This means that an instructor should not give four or five verbal prompts before moving to the demonstration level, though it is acceptable (and desirable) to give a verbal prompt concurrently with the demonstration and physical assistance prompts (Lent & McLean, 1976). The subjects in this study were all familiar with the least-prompts strategy and had used it regularly before the study began.

Data Sheet Formats

Two different forms of the data sheet were used by the subjects with each instructional program. The first format, called the Prompting format, is shown in Figure 1. The Prompting format lists the skill(s) to be taught in the left-hand column. Next to each skill listing are four boxes, marked I, V, D, P, respectively. These initials represent the four prompting levels. The instructor was to mark a plus (+) in the box corresponding to the level of prompting needed to get the student to make a correct response during a trial. For example, in Figure 1, it can be seen that the instructor recorded that the student needed a physical prompt (P) to respond correctly to the first trial of Matching, and a verbal prompt (V) to respond correctly to the first trial of the Kit Assembly program.

Insert Figure 1 about here

The second form of the data sheet, called the Correct/Incorrect format, is shown in Figure 2. This format listed each of the skills next to a numbered grid. This grid was composed of the numbers 0-12, with the zero at the bottom of the grid. These numbers corresponded to the number of trials to be given during a session. On trial 1 of a skill, if the student performed it at the Independent level, the instructor put a circle around the "1" on the data sheet. If the student required any prompting (whether it was a verbal, demonstration or physical prompt), the instructor drew a slash (/) through the "1". This process continued until all trials of both programs had been completed.

Insert Figure 2 about here

Figure 1. Example of the Prompting data sheet format

Figure 2. Example of the Correct/Incorrect data sheet format

Observation Form

Programs	I	V	D	P	I	V	D	P
1. Find the Same (Match)				+				
2. Kit Assembly		+						
3. Find the Same (Match)			+					
4. Kit Assembly		+						
5. Find the Same (Match)				+				
6. Kit Assembly	+							
7. Find the Same (Match)			+					
8. Kit Assembly	+							
9. Find the Same (Match)			+					
10. Kit Assembly	+							
11. Find the Same (Match)			+					
12. Kit Assembly		+						
13. Find the Same (Match)				+				
14. Kit Assembly		+						
15. Find the Same (Match)				+				
16. Kit Assembly	+							
17. Find the Same (Match)				+				
18. Kit Assembly		+						
19. Find the Same (Match)				+				
20. Kit Assembly				+				
21. Find the Same (Match)			+					
22. Kit Assembly	+							
23. Find the Same (Match)		+						
24. Kit Assembly		+						

Date	4-19-83	
Time	10:15	
Instructor	Sam	
Student	Tony	

Instructor: Sam

Student: Tony

Area: Prevocational

Program: Find the Same (Matching)

	~~12~~	12	12	12	12	12	12	12	12	12	12
	~~11~~	11	11	11	11	11	11	11	11	11	11
	~~10~~	10	10	10	10	10	10	10	10	10	10
	⑨	9	9	9	9	9	9	9	9	9	9
	~~8~~	8	8	8	8	8	8	8	8	8	8
	~~7~~	7	7	7	7	7	7	7	7	7	7
	⑥	6	6	6	6	6	6	6	6	6	6
	~~5~~	5	5	5	5	5	5	5	5	5	5
	~~4~~	4	4	4	4	4	4	4	4	4	4
	③	3	3	3	3	3	3	3	3	3	3
	~~2~~	2	2	2	2	2	2	2	2	2	2
	~~1~~	1	1	1	1	1	1	1	1	1	1
	0	0	0	0	0	0	0	0	0	0	0
Date	4-6-83										

Program: Kit Assembly (1-part)

	~~12~~	12	12	12	12	12	12	12	12	12	12
	~~11~~	11	11	11	11	11	11	11	11	11	11
	~~10~~	10	10	10	10	10	10	10	10	10	10
	~~9~~	9	9	9	9	9	9	9	9	9	9
	~~8~~	8	8	8	8	8	8	8	8	8	8
	⑦	7	7	7	7	7	7	7	7	7	7
	⑥	6	6	6	6	6	6	6	6	6	6
	⑤	5	5	5	5	5	5	5	5	5	5
	④	4	4	4	4	4	4	4	4	4	4
	③	3	3	3	3	3	3	3	3	3	3
	②	2	2	2	2	2	2	2	2	2	2
	①	1	1	1	1	1	1	1	1	1	1
	0	0	0	0	0	0	0	0	0	0	0
Date	4-6-83										

In Figure 2, for example, it can be seen that the student performed at the Independent level on trials 3, 6, and 9 of the Matching program, and required prompting, at some level, on all the other trials of this program. On the Kit Assembly task, the student performed at the Independent level on trials 1-7, but needed prompting on the other trials.

It was hypothesized that the Prompting format would result in a more accurate use of the least-prompts strategy than would the Correct/Incorrect format, because the Prompting data sheet had intrinsic reminders to use all levels of prompts in a specific sequence.

Measurement

The observer(s) recorded how accurately the subjects implemented the least-prompts strategy during their sessions. This was accomplished through the use of the observation form illustrated in Figure 3. Using this form, the observer designated the level(s) of prompt the instructor gave and the order in which those prompts were given. This was done by marking numerals in the large boxes (e.g., if the first prompt was a verbal prompt, a "1" was placed in the box marked "V"). The student's response was marked in the corresponding small box (e.g., if the student did not respond correctly to the verbal prompt, a "0" was placed in the small "V" box; if he then responded correctly to the demonstration prompt, a "+" was placed in the small "D" box).

Insert Figure 3 about here

This recording system allowed the observer to determine if the instructor used the strategy correctly (see Trials 1-4 on Figure 3 for examples of correct implementation) or if any of the following errors occurred.

Figure 3. Example of the form used by the observers to record accuracy of using the least-prompts strategy

Observation Form

Programs	I	V	D	P	I	V	D	P
1. Find the Same (Match)	1 +							
2. Kit Assembly	1 0	2 +						
3. Find the Same (Match)	1 0	2 0	3 +					
4. Kit Assembly	1 0	2 0	3 0	4 +				
5. Find the Same (Match)	1 0	3 +	2 0					
6. Kit Assembly	1 0	2 0		3 +				
7. Find the Same (Match)	1 0	2,3,4 0	5 +					
8. Kit Assembly	1 0	2 0	3 +	4 +				
9. Find the Same (Match)	1 0	2 0						
10. Kit Assembly								
11. Find the Same (Match)								
12. Kit Assembly								
13. Find the Same (Match)								
14. Kit Assembly								
15. Find the Same (Match)								
16. Kit Assembly								
17. Find the Same (Match)								
18. Kit Assembly								
19. Find the Same (Match)								
20. Kit Assembly								
21. Find the Same (Match)								
22. Kit Assembly								
23. Find the Same (Match)								
24. Kit Assembly								

Date	4-19-83	
Time	10:15-10:30	
Instructor	Sam	
Student	Tony	
Observer	Laurie	
Reliability		

1) Giving the prompts out of sequence: For example, the instructor might give a demonstration prompt before a verbal prompt. In Figure 3, this type of error was made by the instructor on Trial 5 (row 5 of the observation form).

2) Forgetting to give one or more levels of prompt: For example, the instructor might give a verbal prompt, then a physical assistance prompt. In Figure 3, this type of error was made by the subject on Trial 6 (row 6 of the observation form).

3) Giving repeated prompts at one level rather than immediately progressing to the next level of prompt: For example, the instructor might give repeated verbal prompts before giving a demonstration prompt. This type of error can be seen on Trial 7 of Figure 3.

4) Continuing to prompt after the student made a correct response: For example, the instructor could give a physical prompt after the student had made a correct response to the demonstration prompt. This type of error can be seen on Trial 8 of Figure 3.

5) Discontinuing prompting before the student made a correct response: For example, the instructor might stop prompting (i.e., go on to the next trial) even though the student did not respond correctly to the verbal prompt. This type of error can be seen on Trial 9 of Figure 3.

At the end of a session, the observer left the room and calculated the number of prompting errors made by the instructor. More than one error could be made during a trial. For example, if an instructor gave three verbal prompts, then gave a physical prompt, three errors were counted (two for extra verbal prompts; and one for forgetting the demonstration prompt).

Four measures were then computed for every session. These measures were calculated over the whole session, with no differentiation made between the two programs or students that were taught by the one instructor.

The first variable of interest was the percent of correctly delivered prompts during the session. This was calculated by dividing the number of correctly delivered prompts by the total number of prompts given. The second variable measured was the percent of trials in which a verbal prompt was delivered correctly. This was computed by dividing the number of trials in which a verbal prompt was delivered correctly by the total number of trials in which a verbal prompt should have been given. The third variable was the percent of trials in which a demonstration prompt was given correctly. The fourth was the percent of trials in which a physical assistance prompt was given correctly.

Procedures

The observers engaged in six days of pretraining in the use of the observation form. This was done by having the observers score videotaped sessions. These videotapes were of the instructors who would serve as subjects and showed them teaching the skills that would later be observed during the study. The use of videotapes allowed the observers to review those trials where there was a disagreement between the observers, and determine how such trials should be scored in the future. By the end of the pretraining period, the observers were able to score a session with 80-100% agreement.

During the study, the observer(s) watched the actual sessions rather than relying on videotapes. The instructor was given the appropriate data format and was shown how to use it. Sessions were conducted and observed four days each week. The observer(s) sat approximately six feet from the instructor and student during the sessions.

Experimental Design

An ABA reversal design (Baer, Wolf, & Risley, 1968) was used with Denise and Sam. An AB design was used with Lisa because of time limitations. The order of the data sheet formats was counterbalanced across subjects. This meant that Denise and Lisa used the Correct/Incorrect format first, whereas Sam used the Prompting format first. This counterbalancing was done to determine whether the same results were obtained regardless of the sequence of treatment or whether there might be a sequencing effect.

Reliability

Interobserver reliability measures were taken at least three times in each condition with each instructor. The reliability observer sat apart from the instructor and the primary observer. Data were taken on the number of prompts given, the order of prompting, and the student's performance. Interobserver agreement for the primary observer's recording of teacher accuracy was calculated using the exact-agreement method (Repp, Deitz, Boles, Deitz, & Repp, 1976). An agreement was defined as a trial in which both observers agreed on the level(s) of prompting used, the number of prompts given at each level, the sequence of prompts, and the level at which the student performed correctly. If the observers did not agree on one or more of these categories, a disagreement was scored for that trial. Several disagreements, thus, could be scored on a single trial. For example, on one trial, the primary observer recorded that the instructor gave the student an opportunity to do the skill independently (I = prompt 1), then gave three verbal prompts (V = prompts 2, 3, & 4) before the student responded correctly. The reliability observer, on the other hand, recorded that the instructor gave an independent prompt (I = prompt 1), then only one verbal prompt (V = prompt 2) before the student responded correctly. Two disagreements were recorded because the observers did not agree that verbal prompts 3 and 4 occurred. Reliablity scores for a session were determined by dividing the number of agreements by the number of agreements plus disagreements and converting the quotient to percent.

RESULTS

Reliability

Reliability measures were taken during 10 sessions where Denise used the Correct/Incorrect format. The mean reliability for these sessions was 82% with a range from 54-98%. Reliability measures were taken five times during the Prompting format condition with this subject. The mean reliability in this condition was 89% with a range from 74-96%.

Six sessions of reliability were taken during the Prompting format condition with Sam. The mean reliability for this condition was 81% with a range of 70-87%. In the Correct/Incorrect format condition, a mean reliability of 80% with a range of 48-99% was obtained during seven reliability sessions with Sam.

With Lisa, reliability measures were taken six times in the Correct/Incorrect format condition. A mean reliability of 84%, with a range of 65-98%, was obtained. Reliability measures taken on five of Lisa's sessions during the Prompting condition showed a mean of 92%, with a range of 88-100%.

Prompting Accuracy

Figures 4-6 graphically illustrate the data related to accuracy of using a least-prompts strategy. Denise's data are summarized in Figure

4, Sam's in Figure 5, and Lisa's in Figure 6. On each Figure, the graph in the upper left illustrates the percent of correctly delivered prompts in each session. The graph in the lower left illustrates the percent of trials in which the verbal prompt was delivered correctly. The graph in the upper right shows the percent of trials in which the demonstration prompt was delivered correctly. Lastly, the graph in the lower right shows the percent of trials in which the physical assistance prompt was delivered correctly. The physical assistance graph was not included in Figure 4 because Denise never had an opportunity to give a physical assistance prompt. On the three prompting graphs (Verbal, Demonstration, and Physical assistance), an X was marked if the instructor did not have an opportunity to give that level of prompt during the session. The mean score for each condition is marked on each graph for easy reference.

Insert Figures 4, 5, & 6 about here

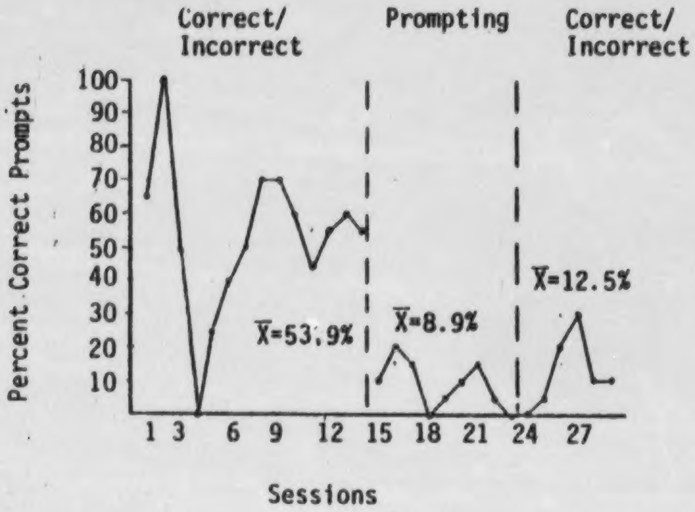
Figure 7 graphically illustrates the percent of independent responses made by the handicapped student(s) during each session. The top graph is a composite of the scores made by Denise's two students. The middle graph is a composite of the scores made by Sam's student on his two programs. The bottom graph represents the scores made by Lisa's student.

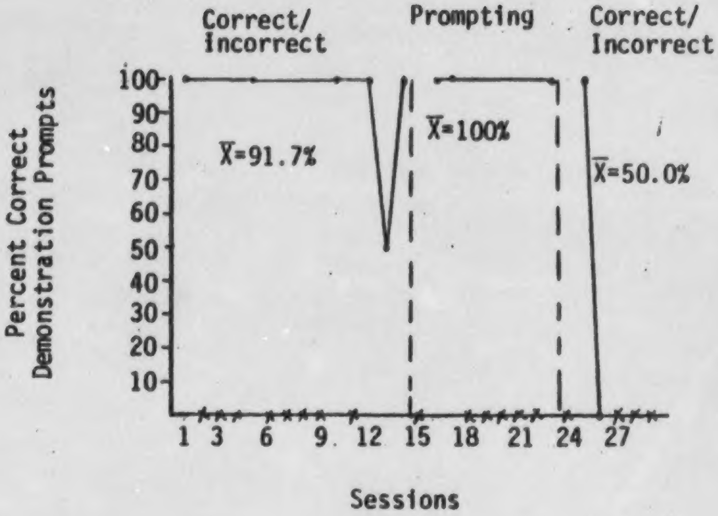
Insert Figure 7 about here

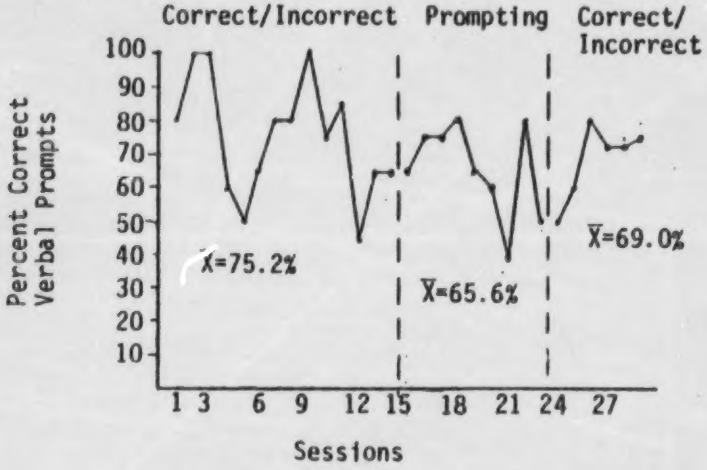
Denise. In Figure 4, it can be seen that Denise showed an immediate and substantial decrease in the percent of correctly delivered prompts when she began using the Prompting format. This depression of accuracy continued throughout the condition. When the Correct/Incorrect format was reintroduced, she improved slightly but did not reestablish the levels achieved earlier with this format.

Figure 4. Denise's accuracy in using the various components of the least-prompts strategy. The upper left graph illustrates present of correctly given prompts. The lower left graph illustrates percent of trials in which the verbal prompt was given correctly. The upper right graph illustrates percent of trials in which the demonstration prompt was given correctly.

Denise



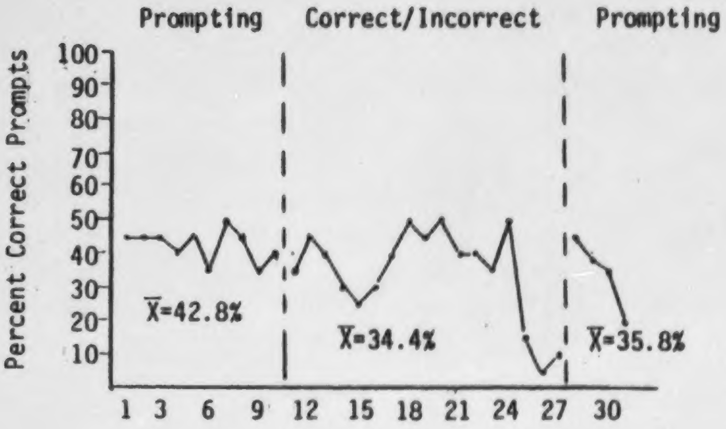


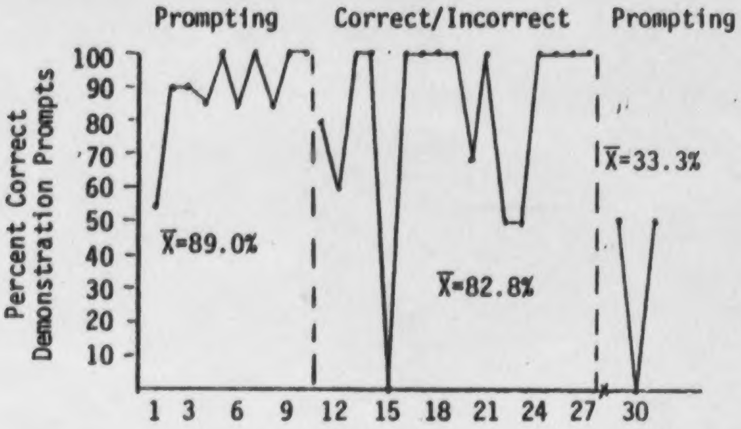


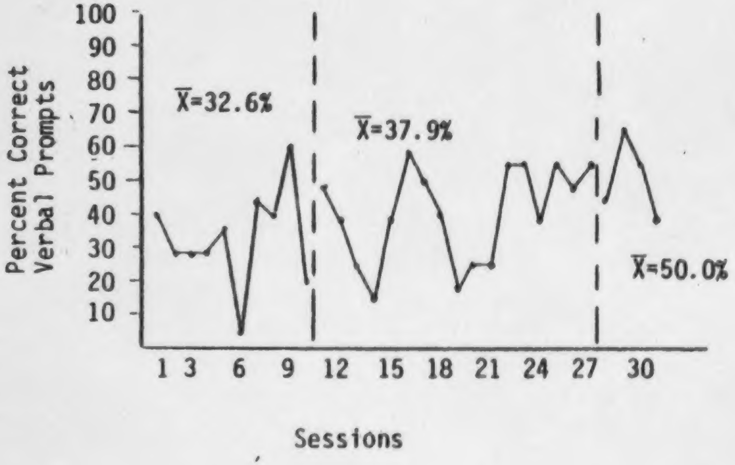
x = no opportunity

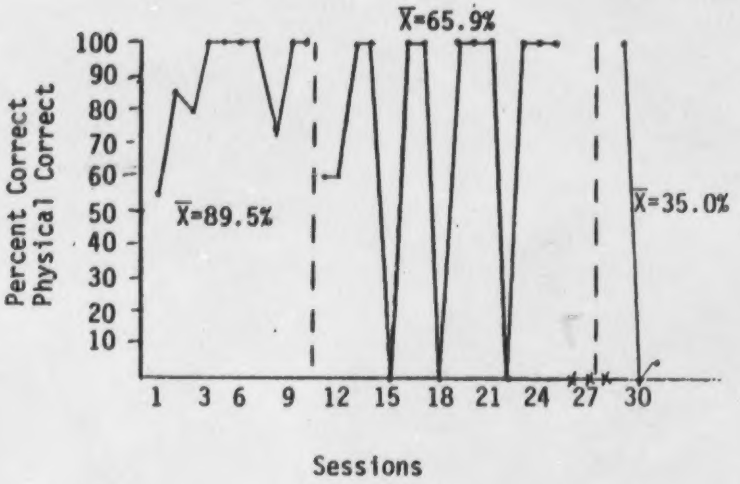
Figure 5. Sam's accuracy in using the various components of the least-prompts strategy. The upper left graph illustrates percent of correctly given prompts. The lower left graph illustrates percent of trials in which the verbal prompt was given correctly. The upper right graph illustrates percent of trials in which the demonstration prompt was given correctly. The lower right graph illustrates percent of trials in which the physical assistance prompt was given correctly.

Sam





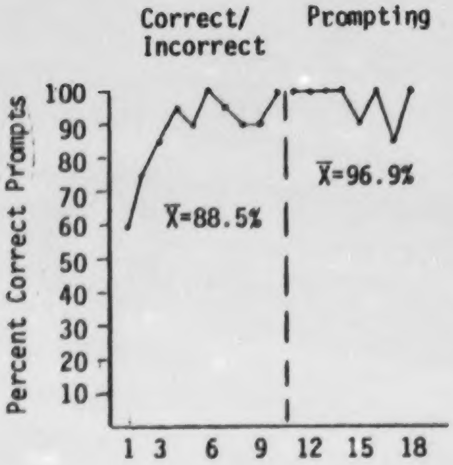


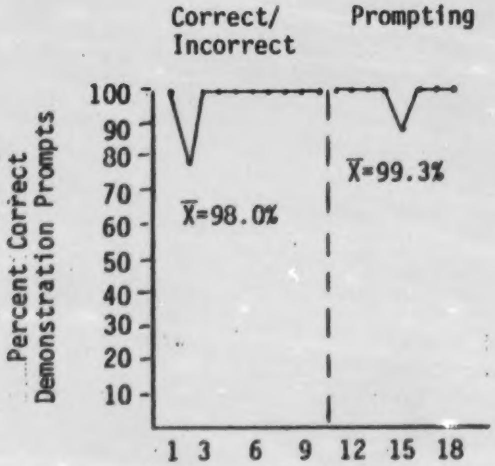


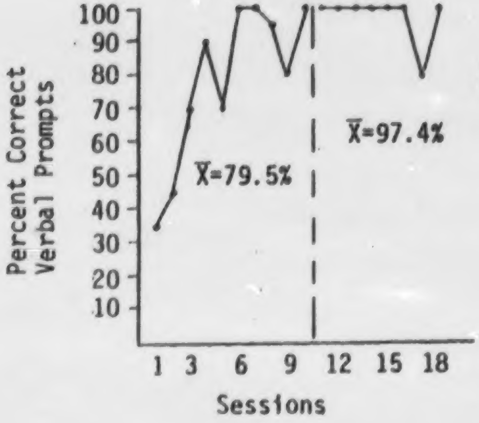
x = no opportunity

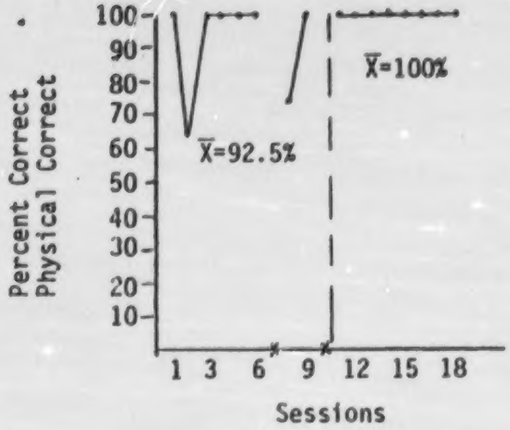
Figure 6. Lisa's accuracy in using the various components of the least-prompts strategy

Lisa



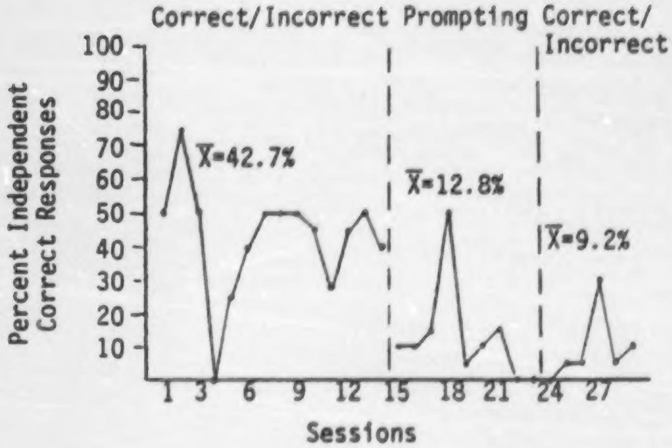




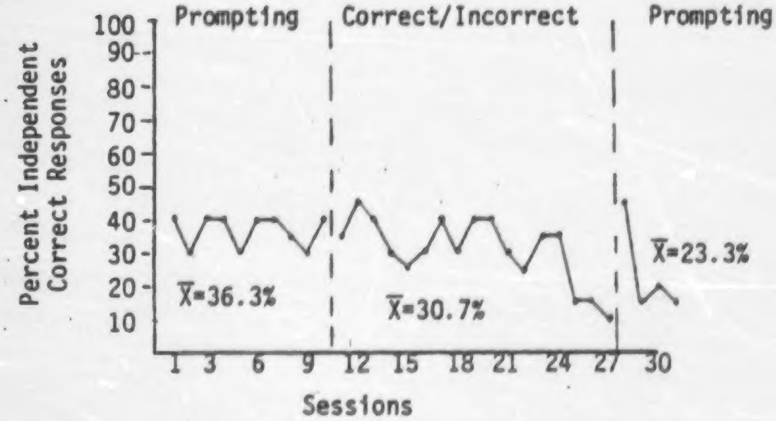


x = no opportunity

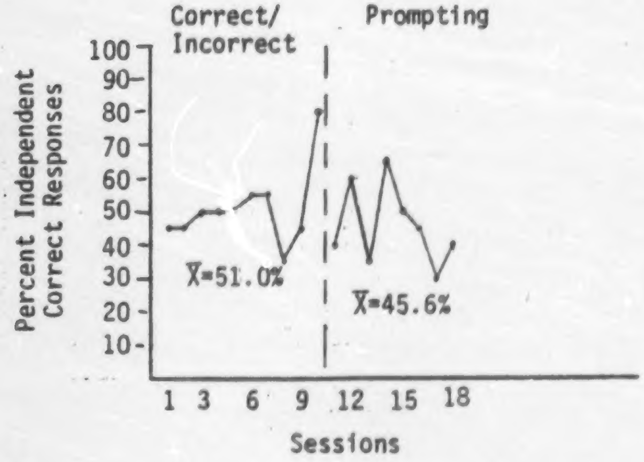
Figure 7. The composite percent correct responses made by the handicapped students while their instructors were using the two data sheet formats



Denise's students



Sam's student



Lisa's student

The verbal prompting graph indicates that Denise gave verbal prompts accurately about 70% of the time across conditions, though the accuracy of verbal prompting decreased somewhat when the Prompting format was introduced.

The demonstration prompting graph indicates that Denise rarely had to use the demonstration prompt, but was generally accurate when she did. Although there are substantial differences in the mean scores across conditions, visual analysis of the data shows that these differences are based on one session.

Looking at Figure 7, it can be seen that Denise's student(s) showed a trend that mirrored the percent of correctly delivered prompts. There was a marked decrease in the student's performance when Denise began using the Prompting format. This decrease in performance was not reversed when the Correct/Incorrect format was reintroduced.

Sam. Sam's data (see Figure 5) indicate that he presented slightly more prompts accurately when the Prompting format was used. The differences, however, were very small.

Sam's verbal prompting graph indicated very little difference between the initial use of the Prompting format and the use of the Correct/Incorrect format. When he returned to the Prompting format, however, there was a slight improvement in his verbal prompting accuracy. In spite of these differences between conditions, it should be noted that 50% or fewer of the trials with verbal prompts were conducted correctly in most sessions and in both conditions.

The demonstration prompting graph showed very accurate use of the demonstration prompt in the initial Prompting format condition and in the Correct/Incorrect format. There was an appreciable drop in accuracy with this prompt, however, when the Prompting format was reintroduced.

The physical assistance prompting graph showed a high and stable performance during the first Prompting condition. Sam's accuracy in giving physical assistance was more variable, and lower when he began using the Correct/Incorrect format. His accuracy with this prompt then decreased even further when he returned to the Prompting format.

Though Sam made errors at all levels of prompting, it appeared that his most serious and consistent problems were at the Verbal level.

Sam's student (see Figure 7) showed a slowly deteriorating performance across conditions. It is interesting to note that this student showed particularly low scores in Sessions 25-27. A similar trend is seen in the percent of prompts presented correctly by Sam on these sessions (Figure 5).

Lisa. Lisa showed a very high percent of correctly delivered prompts in both conditions (see Figure 6). The mean is lower during the Correct/Incorrect format condition, but this appears to be due to errors made during the first three sessions.

Similarly, Lisa made many Verbal prompting errors during the first three sessions of the Correct/Incorrect format, but then achieved and maintained a high level of accuracy in both conditions.

There is virtually no difference in the accuracy of Lisa's demonstration prompts in the two conditions, with scores of 100% being prevalent in both.

A few errors in giving the physical assistance prompt were seen when Lisa was using the Correct/Incorrect format, but none were seen when the Prompting format was used.

Very slight differences in student performance (Figure 7) were seen between the two conditions, with the Correct/Incorrect format being favored.

DISCUSSION

Overall, these data showed a correlation between the accuracy of prompting and the degree of the instructor's experience/education. The instructor with the most training and experience (Lisa) demonstrated the most accurate use of the least-prompts strategy. The instructors with less experience and training made more prompting errors throughout the study. It appeared that the majority of prompting errors occurred during the verbal prompting procedure. All instructors obtained substantially lower scores on their verbal prompts than they did on the demonstration and physical assistance prompts. Analysis of these errors indicated that the most common error was repetition of the verbal cue (often 3 or more times) before giving a demonstration prompt.

Sam and Lisa both were slightly more accurate in using the least-prompts strategy when the Prompting format was used. The difference in scores between conditions, however, was small for these two instructors. Denise, on the other hand, showed a very large decrease in accuracy when she began using the Prompting format. She then was unable to attain her previous high levels of accuracy when she returned to the Correct/Incorrect format. The reason for the drop in accuracy is unclear, though it is possible that the complexity of the Prompting format may have interfered with the prompting process in some way. This instructor repeatedly told the observer that she preferred the Correct/Incorrect format, though she never stated why. The continued inaccuracy of prompting when the Correct/Incorrect format was reintroduced could be the result of an error habit learned during the Prompting condition.

These results could also be a result of temporal variables. This study was done during the last two months of the school year. Both

Denise and Sam expressed that they were very tired of teaching these programs and that they would be glad when school was out. Both these instructors showed some decrement in performance when they entered the second condition, and both showed little or no reversal of performance when the original condition was reinstated. These observations lend some credence to the hypothesis that temporal variables might have had some effect.

A third explanation for the large decrement in performance with Denise might be that there was some misunderstanding about the purpose of the study. At one point during the study, Denise asked the observer when the researchers would collect the student's data from her. She was told that the researchers did not need to see those data. At the debriefing she told the observer that she couldn't understand why she had had to use different data sheet formats if no one was ever going to look at the data she collected. All the instructors mentioned that it was difficult to be observed on a daily basis for two months with no feedback, though only Denise made statements which indicated that this may have affected her performance.

The covariation of the student's performance and the instructor's prompting accuracy for Sam and Denise was very interesting. Initially, it appeared that this covariation was a clear demonstration that errors in prompting had deleterious effects on student learning. It was pointed out, however, by one of the observers that whenever a student did not do well, the instructor had to give more prompts which increased the chances of the instructor making a prompting error. This latter explanation may or may not account for some of the error, but the fact that Lisa's student showed very stable performance levels when the prompts were delivered accurately across conditions and the fact that Denise's students showed a marked decrease in performance on the day the instructor changed data sheets makes it appear that the accuracy of prompting was more likely to influence learning than visa-versa.

In summary, there was some evidence that the Prompting format could help an instructor slightly improve the accuracy of a least-prompts strategy. The difference in accuracy between formats was so small, however, that it is hypothesized that many teachers would prefer to use the simpler format. The hypothesis that the use of the Prompting format would increase accuracy to the degree that it might be perceived as a source of immediate reinforcement was not supported.

Further research needs to be implemented in this area. Specifically, it would be helpful to have a study which investigated whether teaching accuracy was better under data or under no-data conditions. Attention also needs to be directed toward making the data collection process more immediately reinforcing to classroom personnel.

References

Baer, D. M., Wolf, M. M., & Risley, T. R. Some current dimensions of applied behavior analysis. Journal of Applied Behavior Analysis, 1968, 1, 91-97.

Craighead, W. E., Kazdin, A. E., & Mahoney, M. J. Behavior modification: Principles, issues, and applications. Boston: Houghton Mifflin, 1976.

Favell, J. E. The power of positive reinforcement: A handbook of behavior modification. Springfield, IL: Charles C Thomas, 1977.

Hanson, M. J. Teaching your Down's syndrome infant: A guide for parents. Baltimore, MD: University Park Press, 1978.

Haring, N. G., Liberty, K. A., & White, O. W. Rules for data-based strategy decisions in instructional programs: Current research and instructional implications. In W. Sailor, B. Wilcox, & L. Brown (Eds.), Methods of instruction for severely handicapped students. Baltimore, MD: Paul H. Brookes, 1980.

Lent, J. R. Teaching daily living skills, In J. M. Kauffman, & J. S. Payne (Eds.), Mental retardation: Introduction and personal perspectives. Columbus, OH: Charles E. Merrill, 1975.

Lent, J. R., & McLean, B. M. The trainable retarded: The technology of teaching. In N. G. Haring, & R. L. Schiefelbusch (Eds.), Teaching special children. NY: McGraw-Hill, 1976.

Liberty, K. Data dilemma: Response and measurement units for teachers of the severely handicapped. AAESPH Review, 1976, 1(5), 13-31.

Miller, L. K. Principles of everyday behavior analysis. Monterey, CA: Brooks/Cole, 1975.

Mori, A. A., & Masters, L. F. Teaching the severely mentally retarded: Adaptive skills training. Germantown, MD: Aspen Systems, 1980.

Repp, A. C., Deitz, D. E., Boles, S. M., Deitz, S. M., & Repp, C. F. Differences among common methods for calculating interobserver agreement. Journal of Applied Behavior Analysis, 1976, 9(1), 109-113.

Sailor, W., & Guess, D. Severely handicapped students: An instructional design. Dallas, TX: Houghton Mifflin, 1983.

Say it ain't so. Newsletter for the Association for the Severely Handicapped, 1981, 7(9), 1.

Snell, M. E., & Smith, D. D. Intervention strategies. In M. E. Snell (Ed.), Systematic instruction of the moderately and severely handicapped. Columbus, OH: Charles E. Merrill, 1978.

Sternberg, L., & Adams, G. L. Educating severely and profoundly handicapped students. Rockville, MD: Aspen Systems, 1982.

Thompson, T., & Grabowski, J. Behavior modification of the mentally retarded. N.Y.: Oxford University Press, 1972.

Wehman, P., Bates, P., & Renzaglia, A. Characteristics of an appropriate education for severely handicapped students. In P. Wehman & J. Hill (Eds.), Instructional programming for severely handicapped youth: A community integration approach. Richmond, VA: Virginia Commonwealth University, Department of Education, 1980.

White, O. R., & Liberty, K. A. Behavioral assessment and precise educational measurement. In N. G. Haring & R. L. Schiefelbusch (Eds.), Teaching special children. N.Y.: McGraw-Hill, 1976.

Evaluation of Choice and Ready Responses on the Learning of Other Skills

The papers in this section of the manual investigated how responses should be sequenced for optimal learning. The first paper looked at whether allowing students to choose the order in which educational tasks are presented had an effect on the learning of the tasks. The results indicated that allowing such choice had a slight positive effect on learning.

The second paper addressed the issue of whether requiring a ready response (such as hands in lap) before a communication response caused deficits in the initiation of the communication response in other situations. The results of this study were somewhat unclear, but the data seem to indicate that the potential exists for the ready response to interfere with generalization of communication. It may, therefore, be wise to avoid sequencing ready responses and communication responses whenever possible.

Influence of Activity Choice Among Adolescent, Severely, Multiply Handicapped Students

Jennifer Holvoet, Ph.D.

Mike Brewer, M.S. Ed.

Marilyn Mulligan, Ph.D.

Doug Guess, Ed.D.

Ed Helmstetter, M.S. Ed.

Department of Special Education
University of Kansas
Lawrence, Kansas

Perrin Riggs, M.S. Ed.
Department of Education
Kansas Neurological Institute
Topeka, Kansas

1981

This paper was prepared as part of the work pursuant to the "Kansas Individualized Curriculum Sequencing Project" funded as a Handicapped Children's Model Project, U.S. Office of Education (G00-800-1721) from 9-1-80 to 9-1-83.

ABSTRACT

The study evaluated the effects of allowing adolescents with severe handicaps to make limited choices of educational activities. The data show that these students were able to make choices of educational activities when given the opportunity. One activity, assembly of a snap-together plastic model, seemed to be a highly prefered task. This preference was discussed in relation to current trends toward providing age-appropriate classroom and leisure activities. Over several sessions, the students each chose to work on all the educational activities which were offered. However, this variation in task choice sometimes did not occur until the later sessions. Further, it appeared that the choice procedure has a slight facilitative effect on the rate of learning and on correct performance of the educational activities.

Applied behavior analysis techniques, such as contingent reinforcement, shaping, use of criterion measures, etc. have been repeatedly documented as successful methods of teaching students with severe, multiple handicaps. However, these methods have the possible disadvantage of putting most of the control of the environment and responsibility for behavior change in the hands of instructional personnel instead of in the hands of the students. The importance of controlling one's own behavior and the environment has been recognized for many nonhandicapped populations (Holt, 1976; Kozol, 1972). In addition to teaching some selfresponsibility, the literature suggests that allowing students educational choices might also have some positive effects on learning rate and motivation. Several different kinds of choice have been investigated. Generally a group design with yoked controls has been the mode of investigation. This design uses pairs of subjects who are matched according to some criterion, with one subject of each pair being assigned to the experimental group and the other to the control group. During the investigation the experimental subject is studied first and his responses (in this case, the choices he makes) determine which tasks are given to the "yoked" control. For example, when grade school students were allowed to choose between group or individual instruction for the learning of sight-words, the students given the choice learned better than a yoked control group (Berk, 1976). Similarly, when elementary-age students were allowed to choose among five reinforcers (Hockstra, 1978) or to decide how many problems they would have to do correctly to be reinforced (Felixbrod & O'Leary, 1973), higher rates of working and equivalent learning rates were obtained. White (Note 1) investigated the choice of stories (selected from a list of titles) that would be included on a standardized reading comprehension test given to fifth grade students. Some children were allowed no choice, some were allowed partial choice, and others were permitted to choose all the stories. Performance was measured by the score on the standardized test. Results showed no differences in performance between the full- and partial-choice groups, and both choice groups performed significantly better than children in the yoked control group.

An extensive set of studies conducted by investigators at Virginia Polytechnic Institute (Monty & Perlmuter, 1975; Monty, Rosenberger & Perlmuter, 1973; Perlmuter & Monty, 1973; Perlmuter & Monty, 1977) measuring the effects of choice on the learning of adults found that: a) choosing either stimulus or response items enhanced learning, b) if individuals were allowed to choose items, and the experimenter then ignored the choice and had them do other items, performance was poorer than when the student did the task under no-choice conditions, and c) allowing even limited choices was as effective in enhancing learning as allowing the individual to choose all the stimuli or responses, if choice was given early in the session.

Despite all these findings, little systematic effort hs been made in special education, particularly in classrooms for severely handicapped students, to use educational procedures that allow students to exercise choice. However, allowing students to make limited choices about some aspects of their education (e.g. choice of reinforcers, choice of materials, choice of when to do an activity, and possibly even choice of activities) is one method that would permit the teacher to maintain the precision of behavioral teaching techniques, (i.e. the teacher could still use a sequence of prompts task analysis, data analysis) while still providing the opportunity for the students to control certain components of the instructional process.

In one of the few studies done with handicapped persons, Lovitt and Curtis (1969) showed that allowing a 12-year old child with severe emotional problems to set his own criteria on academic tasks resulted in improved performance. This occurred even though the child often chose more stringent requirements than the teacher had previously used. On the other hand, Alexander (1974) demonstrated that allowing adolescents with moderate mental retardation to choose their reinforcers, either for the whole session or for each trial, did not lead to higher levels of performance on discrimination or perseverance tasks.

The present study investigated the impact of choice of activity on the learning of academic and prevocational skills among adolescent students who were severely retarded. In particular, this study addressed: (a) whether the students preferred certain tasks over others; (b) whether the students showed variation, over time, in what they chose or consistently chose the same task; and (c) whether the choice procedure affected acquisition rates and/or mean performance on the tasks.

Method

Subjects

The study used a yoked-control experimental design. Accordingly, it was necessary to select pairs of students with similar learning patterns on a task that was representative of the instructional activities used in the experimental conditions. Four adolescent students (Mary, Flora, Rich and Bill) enrolled in public school classrooms for the severely handicapped participated in this study. All were between 14-17 years of age (mean age = 16) and were classified as severely retarded in terms of both MA/CA ratio and adaptive behavior, using tests such as the WISC-R and the TARC Assessment Guide (Sailor & Mix, 1975). All four of the students consistently followed simple instructions, although each of them had less than 10 word expressive vocabularies. Two of the students (Rich and Mary) regularly engaged in "acting out" and aggressive behaviors and were on programs designed to decelerate these behaviors. Three of the students lived in an institution for the mentally retarded and the other lived at home.

The four students were chosen from a group of 10 severely handicapped adolescents who had been pretested on a prevocational task of counting out ten sticks on a template, gathering them together, fastening them with a rubber band, and putting them in a box. Individual learning curves were calculated for each step of this prevocational task. The Sign Test (Conover, 1971) comparing datum points in the learning curve was used to determine the similarity of a student's performance with every other student on each step of the task. Students with the most similar learning curves for 3 out of 4 steps were selected. Mary and Flora showed the most similar learning patterns, and Rich and Bill showed the next most similar patterns.

Setting

All students were enrolled in self-contained special education classrooms for severely/multiply handicapped (SMH) students. These classrooms were located in a junior high building that primarily served special populations. The students were provided daily instruction on self-help, vocational, mobility, fine motor, gross motor, social and communication tasks. Educational programs in this setting were, as far as possible, functional and chronologically age-appropriate.

The study was conducted in one of the SMH classrooms within the students' school, using a one-to-one instructional format. The student was seated opposite the teacher at a 60 X 66 cm. desk. The training took place during regular class time so other teachers and their students were working in the room at the same time.

Tasks and Instructional Objectives

The classroom teachers were given at a list of 22 tasks and asked to choose ten tasks which seemed suitable for the students participating in the study. The experimenter then picked six of the tasks upon which the teachers agreed for inclusion in the study. These six tasks were arbitrarily divided into two sets, Set A and Set B. The three tasks comprising Set A were:

1) matching capital letters of the alphabet to small letters;
2) putting several stapled notebook papers in a three-ring binder; and
3) assembling a flashlight.

The second set of tasks, Set B, was composed of the following:

1) partial assembly of a SNAP-TITE® plastic model of a car or airplane;
2) filling a stapler and using it; and
3) using wire cutters to trim exposed wire from sections of precut wire.

Complete instructional programs were written for each of the six tasks in accordance with the following instructional objectives:

1. Letter Matching: When shown a capital letter (Y, U, S, O or W) and asked to find the matching small letter from an array of three letters, the student will find the correct small letter and place it next to the capital letter with at least 90% accuracy for three consecutive sessions;

2. Notebook Assembly: When given a closed three-ring binder and a sheaf of 6-10 sheets of stapled notebook paper, and told "Put these papers in the notebook," the student will open the notebook, open the rings, correctly place all sheets of paper on at least 2 of the rings, close the rings, and shut the notebook with at least 90% accuracy for three consecutive sessions;

3. Flashlight Assembly: When given a flashlight body, 2 batteries, and the flashlight top and told "Put the flashlight together," the student will insert both batteries using the correct orientation and screw the top on far enough for the flashlight to light with at least 90% accuracy for three consecutive sessions;

4. Model Assembly: When handed each piece of a plastic airplane or car model one at a time and told "Put the (wing) on your model," the student will place the piece on the model in the correct spot and in the correct orientation within 10 seconds of the cue with 90-100% accuracy for three consecutive sessions;

5. Filling and Using a Stapler: When given a large office stapler, two sheets of scrap paper and partially filled box of staples and the instruction "Fill the stapler and staple these pages together," the student will open the stapler, open the box of staples, take out a segment of staples, place them in the correct orientation in the stapler, shut the stapler, and staple two pieces of paper together with at least 90% accuracy for three consecutive sessions;

6. Wire Cutting: When given a set of wire cutters, a box and an insulated wire with 1-2 inches of exposed wire at each end, and a demonstration of how to cut off the exposed wire, the student will hold the wire over the box, open the wire cutters, insert the exposed wire, and close the cutters with enough force to cut the wire, leaving no more than 1/8 inch of exposed wire showing. This will be done with at least 90% accuracy for three consecutive sessions.

Experimental Design and Procedure

This study used a counterbalanced design across "yoked" students. One student in each yoked pair ("Choice" student) was allowed to choose among the tasks in one set (e.g., set A). Whatever tasks were chosen by the "Choice" student were taught, not only to that student, but also to a yoked control student. The "Choice" student was presented with the materials from two of the three tasks and told "Here's the ----and here's the ----. Which one do you want to do?" When the student had chosen one of the tasks (either through touching the materials or eye contact with the materials), the student was given several trials of the task. Rich and Bill had only five trials of each task, whereas Mary and Flora had ten trials of each task. This difference was due to the slowness with which Bill executed tasks. This process of being presented with two sets of materials, being asked to select one and being trained on the selected set, was then repeated twice more. This allowed every task in the set to be paired with every other task. Care was taken to insure that the same sequence of choice pairs did not occur in consecutive sessions and that the choice items were randomly positioned to control for position preferences.

When the "Choice" student had completed all three of the chosen tasks, the yoked control was taught. The yoked control student was not shown the choice items, but was simply presented with routine training on the tasks selected by the Choice student.

To illustrate the whole procedure, Mary was allowed to choose between the tasks of matching letters, and assembling the flashlight. She chose to do the flashlight task and was given 10 trials on this task. Then she was asked to choose between matching letters and putting paper in a notebook. She chose to match letters and was given 10 trials of training on this task. Lastly, she was given the choice of assembling the flashlight or putting paper in the notebook and she chose to assemble the flashlight again. She was given another 10 trials of training on the flashlight task. Mary left, and Flora arrived and was asked to do 10 trials of the flashlight task, 10 trials of the matching task, and 10 more trials of the flashlight task, in the same order of the training that had been done by Mary (the "Choice" subject).

Note that on some days a student might select not to do a task at all. In the example given, neither student practiced putting paper in the notebook during the session. However, all students chose to do all of the tasks over several sessions.

The Control student and the "Choice" student were trained consecutively in afternoon sessions on the same day. Each session lasted approximately one hour.

When ten sessions had been completed, the roles were reversed using a new set of tasks (e.g., Set B). The student who had served as a yoked control became the "Choice" subject, and visa versa. The training and choice procedures were conducted in the same way as they had been in the previous condition. It should be noted that Mary and Flora first encountered Set A tasks and then experienced Set B during the reversal. Rich and Bill were taught Set B tasks first, then were changed to Set A. This counterbalancing was done to reduce possible order effects between sets of tasks.

All tasks were taught using standard prompting procedures with social and tangible reinforcers being given for correct performance. Yoked subjects received equivalent amounts of tangible reinforcers to those received by the "Choice" subjects during sessions conducted on the same day.

Reliability

Interobserver reliability was assessed at least once in every condition for each student. The same number of reliability sessions could not be held for every task since it could not be known at the beginning of a session which tasks the student would choose, and reliability observers were only available intermittently. To obtain reliability scores, the observer sat near the experimenter where she/he could clearly see the student's response, but where the experimenter's data sheet could not be seen. Reliability was calculated for each task separately by dividing the number of trials on which the two observers agreed by the total number of trials observed for that program during the session X 100. Several different staff members served as reliability observers.

Results

Reliability data for the students' performance on the various tasks ranged from 78-100%, overall, and averaged 90% for Mary/Flora and 97% for Rich/Bill.

Several questions were addressed on this study: (a) Whether the students preferred certain tasks over others; (b) Whether the students showed variation, over time, in what they chose or consistently chose the same task; and (c) Whether the choice procedure affected acquisition rates and/or mean performance on the tasks.

In order to answer the first question, data were taken on the number of times a certain task was chosen in a given choice set. The results are shown in Table 1. These data show, for example, when Mary was given the choice of the letter matching task vs. the notebook task, she chose the letter matching task seven times and the notebook task three times. When Rich was given the choice of these two activities, the letter matching task was chosen four times and the notebook task was chosen six times. It can be seen that the model assembly task was chosen as often, or more often, by all students regardless of the task with which it was paired. Other preferences appeared to depend upon the tasks in the choice pair. Ancedotal evidence from the experimenters indicated that these students were able to make choices independently.

Insert Table 1 about here

Table 1

Frequency of Task Choices Made By Students

Student Dyad	Items in Choice Set Item 1	Item 2	Frequency with which Item 1 was chosen	Frequency with which Item 2 was chosen
Mary*/Flora	Letter Match	Notebook	7	3
Mary*/Flora	Flashlight	Notebook	7	3
Mary*/Flora	Letter Match	Flashlight	5	5
Rich*/Bill	Letter Match	Notebook	4	6
Rich*/Bill	Flashlight	Notebook	7	3
Rich*/Bill	Letter Match	Flashlight	6	4
Mary/Flora*	Wire Cutting	Stapler	5	1
Mary/Flora*	Stapler	Model Assembly	3	3
Mary/Flora*	Wire Cutting	Model Assembly	1	5
Rich/Bill*	Wire Cutting	Stapler	6	4
Rich/Bill*	Stapler	Model Assembly	4	6
Rich/Bill*	Wire Cutting	Model Assembly	2	8

*indicates the person choosing

Only one student (Bill) required additional verbal prompting to get him to make the choice and he required this only during the first two sessions.

The pattern of student choice across time is shown in Table 2. All

Insert Table 2 about here

students chose all tasks at least once. It can be seen that the students generally varied their choices somewhat randomly across time. There are, however, two interesting exceptions to this. In the tasks sets, flashlight vs. notebook and letter match vs. notebook, Mary did not choose to do the notebook at all during the first five sessions, but selected to do it at least once in each of the last five sessions. Similarly, Bill had three consecutive sessions (4, 5, 6) in which he did not choose to do the stapler task.

The effect of the choice procedure on learning was addressed by comparing the performance of the yoked pairs on the six tasks. Mean percent correct performance on each task was computed by dividing the number of steps performed correctly by the number of possible steps. Rate of learning was determined by calculating the slope of the least-squares best fit lines across all sessions for each task. These results are shown in Tables 3 and 4. In these tables, the results for all six tasks in a condition were averaged to obtain the data labeled "All Tasks". Similarly, the results for the three tasks making up Set A were averaged to obtain the data labeled "Tasks in Set A." The tasks were also analyzed individually and these data are labeled with the name of the task (e.g. Notebook assembly).

Insert Tables 3 & 4 about here

Table 2

Order of the Tasks Chosen by the Different Students Across Sessions

	Sessions									
Students	1	2	3	4	5	6	7	8	9	10
Mary	F F L	L F L	F F L	L F L	F L F	F F N	N L L	F F N	N L N	L N L
Rich	L F F	N N L	L N N	F F N	F L L	F L F	N L N	L F F	F L N	N F L
Flora	M W M	M S W	W M S	W M M	M W S	M W S				
Bill	M W S	M S M	M W S	M W M	W M M	M W W	M S W	M S M	S M S	S S M

Legend:
F = Flashlight assembly
L = Letter matching
N = Notebook assembly
M = Model assembly
W = Wire cutting
S = Filling and using stapler

Note: The three tasks listed within a session are shown in sequential temporal order, with the first task chosen being listed on the left.

Table 3

Mean Percent Correct Performance on Tasks
Under Choice and Non-Choice Conditions

Tasks/Subjects	Choice Condition	Non-Choice Condition
All Tasks Combined		
Mary/Flora	59%	58%
Rich/Bill	66%	48%
Both dyads	63%	53%
Tasks in Set A		
Mary*/Flora	64%	53%
Rich*/Bill	61%	51%
Both dyads	62%	52%
Tasks in Set B		
Mary/Flora*	55%	63%
Rich/Bill*	72%	46%
Both dyads	63%	54%
Notebook Assembly		
Mary*/Flora	72%	67%
Rich*/Bill	45%	74%
Both dyads	59%	71%
Flashlight Assembly		
Mary*/Flora	59%	59%
Rich*/Bill	65%	74%
Both dyads	62%	67%
Letter Matching		
Mary*/Flora	60%	33%
Rich*/Bill	72%	04%
Both dyads	66%	19%
Model Assembly		
Mary/Flora*	59%	72%
Rich/Bill*	66%	51%
Both dyads	63%	65%
Filling and Using a Stapler		
Mary/Flora*	47%	58%
Rich/Bill*	87%	37%
Both Dyads	67%	48%
Wire Cutting		
Mary/Flora*	59%	58%
Rich/Bill*	67%	49%
Both dyads	61%	54%

Table 4

Learning Slopes for Tasks Under Choice and Non-Choice Conditions

	Choice	Non-Choice
All Tasks		
Mary/Flora	3.1	1.2
Rich/Bill	2.4	1.8
Both dyads	2.7	1.5
Tasks in Set A		
Mary/Flora	4.3	-0.4
Rich/Bill	3.2	0.5
Both dyads	3.8	0.1
Tasks in Set B		
Mary/Flora	1.9	2.5
Rich/Bill	3.6	3.1
Both dyads	2.8	2.8
Notebook Assembly		
Mary/Flora	6.6	-0.7
Rich/Bill	-0.9	1.7
Both dyads	2.8	0.8
Flashlight Assembly		
Mary/Flora	2.6	-0.5
Rich/Bill	3.2	0.8
Both dyads	2.9	0.1
Letter Matching		
Mary/Flora	3.7	0.8
Rich/Bill	7.3	-0.9
Both dyads	5.5	0.1
Model Assembly		
Mary/Flora	2.2	4.0
Rich/Bill	3.8	-6.6
Both dyads	3.5	1.3
Wire Cutting		
Mary/Flora	4.0	5.1
Rich/Bill	-3.1	10.3
Both dyads	0.5	7.5
Filling and Using Stapler		
Mary/Flora	-0.4	-1.6
Rich/Bill	3.9	5.5
Both dyads	4.5	4.7

When the data for all six tasks are averaged, (All Tasks) it can be seen that a higher mean performance and faster learning average is obtained on the tasks under the choice condition, although the differences are nonsignificant. This is true for both dyads (a dyad consisting of an experimental subject and a yoked control). Similarly when the results are averaged for the tasks in Set A, the choice condition results in somewhat better performance averages and steeper learning slope means for both dyads. For the tasks in Set B, however, only the Rich/Bill dyad shows a higher mean and steeper average slope under the choice condition. The dyad of Mary and Flora shows higher levels and faster average learning under the non-choice condition on this set of tasks.

Looking at the tasks individually, it appears that the choice procedure resulted in higher mean performance on three of the tasks (notebook assembly, letter matching, and wire cutting) for the Mary/Flora dyad; and on four of the tasks (letter matching, model assembly, filling and using a stapler, and wire cutting) for the Rich/Bill dyad. In terms of the rate of learning, faster rates occurred under the choice condition for four tasks in the Mary/Flora dyad and three tasks for the Rich/Bill dyad. Letter matching was the only task which seemed to be consistently enhanced, both in terms of mean performance and learning rate, by the choice procedure. It should be noted that, overall, Mary and Rich's learning was better and faster than that of their yoked controls under both the Choice and the Non-choice conditions.

Although there was no matching done between the tasks in Set A and Set B, two subjects chose from Set A and two subjects chose from Set B. Therefore, some further information can be gained through a comparison of the performance of each subject under the Choice condition to his/her performance in the Non-choice condition. This intra-subject comparison, found in Table 5, shows that all subjects learned slightly better and faster under the choice condition.

In summary, these data show that these students with severe multiple handicaps were able to make choices concerning educational activities. One activity, model assembly, seemed to be a highly preferred task for the students. Over several sessions, the students each chose to work on all tasks. However, this variation in task choice sometimes did not occur until the later sessions. It appears the choice procedure has a very slight facilitative effect on the rate of learning and on correct performance of the tasks with these students.

Discussion

The results of this study indicate that these students with relatively severe mental handicap were able to make structured choices about their education. Following the students to make choices of activities did not appear to have any negative impact upon their learning. In some cases, the rate of learning and the number of correct responses actually increased under the choice condition. Though improvement in learning under the choice condition was less than might have been predicted from the bulk of the literature this might be accounted for by several factors. First, it is possible that the difference between the two conditions might have been larger if the non-choice subjects had been shown the two sets of materials (as were the choice subjects) and the teacher then made the choice. Most choice studies have used this method to emphasize that the subject is not being allowed to make the choice. However, it was felt that this method was not a fair comparison of whether choice procedures would be superior to current classroom practices. Secondly, it is possible that the initial matching of the students was not suitable, since Rick and Mary consistently outperformed their partners hastly, it may be that adolescents who are retarded are less sensitive to the choice procedures either inherently or because they have a long history of not being allowed choice, both at home and at school.

Also of interest was the fact that the students seemed to show preference for the model assembly task. This is a leisure task enjoyed by many non-handicapped adolescents. The findings suggest that these severely handicapped students also found this task enjoyable, at least compared to the other available tasks. This finding lends some validation to the statements of several 5MH educators (Brown, Branston, Hamre-Nietupski, Pumpian, Certo & Gruenwald, 1979; Holvoet, Guess, Mulligan & Brown, 1980; Wehman, Schleinen & Kiernan, 1980) that every effort should be made to find choronologically age-appropriate leisure tasks for this population.

Although these results must be interpreted cautiously in view of the number of subjects and the small differences in learning between the condition, the study appears to support the notion that structured "choice" is as effective as current methods and thus might be a valuable adjunct to the traditional classroom techniques used with severely handicapped students. Certainly such choice can be given without sacrificing precision of teaching methods or data collection if the teacher is willing to specify how each available task should be taught. Choice procedures also need not obviate teaching to criterion, since the student could be allowed to choose from different activities which teach the same skill. Even if this were not done, the student could be presented with the same choices over several sessions which should eventually allow the student to reach criterion on each selected task. In addition to the direct effect which choice may have upon learning of tasks, it may have more far reaching effects in that allowing the students to make controlled choices in the educational environment may make them better able to handle choices they will need to make in other environments. Furthermore, the technique of allowing the students some choices in the educational environment makes the classroom a place where the students are treated more as equal partners in the educational experience than as vessels to be filled with whatever the teacher deems best for them.

Reference Notes

1. White, A. M. The effects of choice upon paired-associate learning and reading comprehension in children. 1974 undergraduate honors thesis, Virginia Polytechnic Institute and State University. As cited in Perlmuter, L. C. & Monty, R. A. The importance of perceived control: Fact or fantasy. American Scientist, 1977, 65, 759-765.

References

Alexander, D. Comparison for mental retardates of non-choice, initial-choice, and idiosyncratic-choice reward strategies. Psychological Reports, 1974, 35, 135-145.

Berk, R. A. Effects of choice of instructional methods on verbal learning tasks. Psychological Reports, 1976, 38, 867-870.

Brown, L., Branston, M., Hamre-Nietupski, S., Pumpian, I., Certo, N., and Gruenwald, L. A strategy for developing chronological age-appropriate skills for severely handicapped adolescents and young adults. Journal of Special Education, 1979, 13, 81-90.

Conover, W. J. Practical nonparametric statistics. New York: John Wiley & Sons, 1971.

Felixbrod, J. J., & O'Leary, K. D. Effects of reinforcement on children's academic behavior as a function of self-determined and externally imposed contingencies. Journal of Applied Behavior Analysis, 1973, 6, 241-250.

Hockstra, C. M. The effects of choice of consequences and procedures on preschool children's rate of working arithmetic problems. Unpublished doctoral dissertation, Washington State University, 1978. In Dissertation Abstracts International, 1979, 39(8-B), 4033-4034.

Holt, J. C. Instead of education: Ways to help people do things better. N.Y.: Dutton, 1976.

Holvoet, J., Guess, D., Mulligan, M. & Brown, F. The individualized curriculum sequencing model (II): A teaching strategy for severely handicapped students. Journal of the Association for Severely Handicapped, 1980, 5, 337-351.

Kozol, J. Free schools: A time for candor. Saturday Review, 1972, 55, 51-54.

Lovitt, T. C., & Curtiss, K. Academic responses rate as a function of teacher and self-imposed contingencies. Journal of Applied Behavior Analysis, 1969, 2, 49-53.

Monty, R. A., & Perlmuter, L. C. Persistence of the effects of choice on paired associate learning. Memory and Cognition, 1975, 3, 183-187.

Monty, R. A., Rosenberger, M. A., & Perlmuter, L. C. Amount and locus of choice as sources of motivation in paired-associate learning. Journal of Experiemental Psychology, 1973, 97, 16-21.

Perlmuter, L. C., & Monty, R. A. Effect of choice of stimulus on paired-associated learning. Journal of Experimental Psychology, 1973, 99, 120-123.

Perlmuter, L. C. & Monty, R. A. The importance of perceived control: Fact or fantasy. *American Scientist*, 1977, *65*, 759-765.

Perlmuter, L. C., Monty, R. A., & Cross, P. M. Choice as a disrupter of performance in paired-associate learning. *Journal of Experimental Psychology*, 1974, *102*, 170-172.

Sailor, W., & Mix, B. J. *The TARC Assessment Guide*. Lawrence, KS: H & H Enterprises, Inc., 1975.

Wehman, P., Schleinen, S., and Kiernan, J. Age appropriate recreation programs for severely handicapped youth and adults. *Journal of the Association for Severely Handicapped*, 1980, *5*, 395-408.